High weak order discretization schemes for stochastic differential equation

Clement Rey

Cermics, École Nationale des Ponts et Chaussées

Université Paris-est Marne-la-Vallée

École doctorale MSTIC

Étude et modélisation d'équations différentielles stochastiques

Clément Rey

Thèse présentée pour obtenir le grade de

Docteur de l'Université Paris-Est

Soutenue publiquement le 4 décembre 2015 devant le jury composé de

Directeurs de thèse :	
Aurélien Alfonsi	**(Cermics-ENPC)**
Vlad Bally	**(LAMA-Université Marne-la-Vallée)**
Rapporteurs :	
Arnaud Gloter	**(LaMME-Université d'Évry)**
Denis Talay	**(Inria Sophia-Antipolis)**
Éxaminateurs :	
Emmanuel Gobet	**(CMAP-École Polytechnique)**
Gilles Pagès	**(LPMA-Université Pierre et Marie Curie)**

Remerciements

Avant tout, je souhaite exprimer ma profonde gratitude à l'égard mes deux directeurs de thèse Aurélien Alfonsi et Vlad Bally. Travailler à leur côté pendant ces trois années a été un privilège. Je tiens particulièrement à souligner leur disponibilité ainsi que la bonne ambiance qu'ils faisaient régner lors de nos nombreuses réunions.

Je tiens également à remercier Jean-François Delmas, le directeur du Cermics, et de façon plus générale l'ensemble des membres du Cermics qui ont rendu ces trois années très agréables.

Dans un second temps, j'aimerais remercier tout particulièrement les membres du jury. Je remercie sincèrement Denis Talay pour avoir accepté de rapporter mes travaux ainsi que pour nos échanges toujours agréables lors de divers séminaires. Je remercie également Arnaud Gloter pour son rôle de rapporteur mais aussi parce qu'il a été l'un des mes très bons enseignants lors de mes études ultérieures. J'adresse aussi mes remerciements à Gilles Pagès pour avoir accepté de faire partie du jury mais aussi pour l'opportunité qu'il m'offre de continuer dans la voie de la recherche universitaire. J'ai hâte de pouvoir travailler au sein de son équipe. Enfin je remercie tout particulièrement Emmanuel Gobet, le dernier membre de mon jury. Il a été mon premier professeur de Calcul Stochastique et m'a donné l'envie de travailler dans ce domaine. De plus il m'a toujours conseillé dans mon parcours, et ceci également bien après avoir été mon enseignant.

Je tiens aussi à remercier Ahmed Kebaier avec qui nous avons avons travaillé pendant une grande majorité de ma thèse. Ces séances de travail mais aussi toutes nos discussions ont été un plaisir pour moi.

De façon plus générale, je remercie l'ensemble des personnes que j'ai côtoyées pendant ces trois années, notamment les membres du Cermics du LAMA ou de l'école doctorale MSTIC. Je ne citerai pas l'ensemble des professeurs, post doc, doctorants, stagiaires ou plus généralement chercheurs car j'en oublierais forcément mais je n'oublierai pas les bons moments que j'ai passé avec vous. Enfin je tiens à adresser mes sincères remerciements à Sylvie Cach et Isabelle Simunic qui, en plus d'être très sympathiques, effectuent un travail remarquable et ont toujours résolu mes problèmes administratifs avec une étonnante facilité.

La suite de mes remerciements s'adressent à l'ensemble de mes proches.

Je souhaite simplement remercier mes parents, ma soeur Margaux et mon frère Benoît ainsi que l'ensemble des membres de ma famille pour l'infinité de choses qu'ils ont faites pour moi au cours de ma vie. Ils sont la raison principale de cet accomplissement et je ne peux exprimer en quelques lignes la gratitude que j'ai pour eux.

Je veux aussi remercier l'ensemble de mes amis de Pau, Grenoble, Paris ou d'ailleurs. Ils représentent une part très importante de ma vie et je ne me lasserai jamais des moments passés ensemble. Je compte bien célébrer la fin de ces trois années avec eux. Je veux tout particulièrement remercier mes amis les plus anciens Ben, Erik, Gogo, Gracianne, JD, Laura, Lolo et Max pour leur soutien indéfectible et pour bien d'autres raisons. Je pense également à tous ceux que j'ai pu rencontrer plus récemment Antoine, Claire, Hugo, PP, Thibault, Will, les anciens de l'Ensimag ou encore à ceux que j'ai rencontrés à Paris.

Enfin je suis infiniment reconnaissant envers Sonia qui partage ma vie et qui a elle aussi vécu l'experience de la thèse. Son soutien quotidien m'a permis de traverser ces trois années avec davantage de motivation mais surtout en étant très heureux.

Summary

The development of technology and computer science in the last decades, has led the emergence of numerical methods for the approximation of Stochastic Differential Equations (SDE) and for the estimation of their parameters. This thesis treats both of these two aspects. In particular, we study the effectiveness of those methods. The first part will be devoted to SDE's approximation by numerical schemes while the second part will deal with the estimation of the parameters of the Wishart process.

First, we focus on approximation schemes for SDE's. We will treat schemes which are defined on a time grid with size n. We say that the scheme X^n converges weakly to the diffusion X, with order $h \in \mathbb{N}$, if for every $T > 0$, $|\mathbb{E}[f(X_T) - f(X_T^n)]| \leqslant C_f/h^n$. Until now, except in some particular cases (Euler and Victoir Ninomiya schemes), researches on this topic require that C_f depends on the supremum norm of f as well as its derivatives. In other words $C_f = C\sum_{|\alpha|\leqslant q} \|\partial_\alpha f\|_\infty$. Our goal is to show that, if the scheme converges weakly with order h for such C_f, then, under non degeneracy and regularity assumptions, we can obtain the same result with $C_f = C\|f\|_\infty$. We are thus able to estimate $\mathbb{E}[f(X_T)]$ for a bounded and measurable function f. We will say that the scheme converges for the total variation distance, with rate h. We will also prove that the density of X_T^n and its derivatives converge toward the ones of X_T. The proof of those results relies on a variant of the Malliavin calculus based on the noise of the random variable involved in the scheme. The great benefit of our approach is that it does not treat the case of a particular scheme and it can be used for many schemes. For instance, our result applies to both Euler ($h = 1$) and Ninomiya Victoir ($h = 2$) schemes. Furthermore, the random variables used in this set of schemes do not have a particular distribution law but belong to a set of laws. This leads to consider our result as an invariance principle as well. Finally, we will also illustrate this result for a third weak order scheme for one dimensional SDE's.

The second part of this thesis deals with the topic of SDE's parameter estimation. More particularly, we will study the Maximum Likelihood Estimator (MLE) of the parameters that appear in the matrix model of Wishart. This process is the multi-dimensional version of the Cox Ingersoll Ross (CIR) process. Its specificity relies on the square root term which appears in the diffusion coefficient. Using those processes, it is possible to generalize the Heston model for the case of a local covariance. This thesis provides the calculation of the EMV of the parameters of the Wishart process. It also gives the speed of convergence and the limit laws for the ergodic cases and for some non-ergodic case. In order to obtain those results, we will use various methods, namely: the ergodic theorems, time change methods or the study of the joint Laplace transform of the Wishart process together with its average process. Moreover, in this latter study, we extend the domain of definition of this joint Laplace transform.

Résumé

Durant les dernières décennies, l'essor des moyens technologiques et particulièrement informatiques a permis l'émergence de la mise en œuvre de méthodes numériques pour l'approximation d'Équations Différentielles Stochastiques (EDS) ainsi que pour l'estimation de leurs paramètres. Cette thèse aborde ces deux aspects et s'intéresse plus spécifiquement à l'efficacité de ces méthodes. La première partie sera consacrée à l'approximation d'EDS par schéma numérique tandis que la deuxième partie traite l'estimation de paramètres.

Dans un premier temps, nous étudions des schémas d'approximation pour les EDSs. On suppose que ces schémas sont définis sur une grille de temps de taille n. On dira que le schéma X^n converge faiblement vers la diffusion X avec ordre $h \in \mathbb{N}$ si pour tout $T > 0$, $|\mathbb{E}[f(X_T) - f(X^n_T)]| \leqslant C_f/n^h$. Jusqu'à maintenant, sauf dans certains cas particulier (schémas d'Euler et de Ninomiya Victoir), les recherches sur le sujet imposent que C_f dépende de la norme infini de f mais aussi de ses dérivées. En d'autres termes $C_f = C\sum_{|\alpha|\leqslant q} \|\partial_\alpha f\|_\infty$. Notre objectif est de montrer que si le schéma converge faiblement avec ordre h pour un tel C_f, alors, sous des hypothèses de non dégénérescence et de régularité des coefficients, on peut obtenir le même résultat avec $C_f = C\|f\|_\infty$. Ainsi, on prouve qu'il est possible d'estimer $\mathbb{E}[f(X_T)]$ pour f mesurable et bornée. On dit alors que le schéma converge en variation totale vers la diffusion avec ordre h. On prouve aussi qu'il est possible d'approximer la densité de X_T et ses dérivées par celle X^n_T. Afin d'obtenir ce résultat, nous emploierons une méthode de calcul de Malliavin adaptatif basée sur les variables aléatoires utilisées dans le schéma. L'intérêt de notre approche repose sur le fait que l'on ne traite pas le cas d'un schéma particulier. Ainsi notre résultat s'applique aussi bien aux schémas d'Euler ($h = 1$) que de Ninomiya Victoir ($h = 2$) mais aussi à un ensemble générique de schémas. De plus les variables aléatoires utilisées dans le schéma n'ont pas de lois de probabilité imposées mais appartiennent à un ensemble de lois ce qui conduit à considérer notre résultat comme un principe d'invariance. On illustrera également ce résultat dans le cas d'un schéma d'ordre 3 pour les EDSs unidimensionnelles.

La deuxième partie de cette thèse traite le sujet de l'estimation des paramètres d'une EDS. Ici, on va se placer dans le cas particulier de l'Estimateur du Maximum de Vraisemblance (EMV) des paramètres qui apparaissent dans le modèle matriciel de Wishart. Ce processus est la version multi-dimensionnelle du processus de Cox Ingersoll Ross (CIR) et a pour particularité la présence de la fonction racine carrée dans le coefficient de diffusion. Ainsi ce modèle permet de généraliser le modèle d'Heston au cas d'une covariance locale. Dans cette thèse nous construisons l'EMV des paramètres du Wishart. On donne également la vitesse de convergence et la loi limite pour le cas ergodique ainsi que pour certains cas non ergodiques. Afin de prouver ces convergences, nous emploierons diverses méthodes, en l'occurrence : les théorèmes ergodiques, des méthodes de changement de temps, ou l'étude de la transformée de Laplace jointe du Wishart et de sa moyenne. De plus, dans dernière cette étude, on étend le domaine de définition de cette transformée jointe.

Contents

Notations

General Notations

$\mathbb{N}$: Set of Integers.
$\mathbb{N}^* = \mathbb{N} \setminus \{0\}$.
$\mathbb{R}$: Set of Real numbers.
$\mathbb{R}^* = \mathbb{R} \setminus \{0\}$.
$B_R(x_0) = \{x \in \mathbb{R}^d, |x - x_0| < R\}$, $x_0 \in \mathbb{R}^d, R \geqslant 0$.
$\overline{B}_R(x_0) = \{x \in \mathbb{R}^d, |x - x_0| \leqslant R\}$, $x_0 \in \mathbb{R}^d, R \geqslant 0$.
$\mathcal{M}_d$: Set of real d-square matrices.
$\mathcal{S}_d^+ \subset \mathcal{M}_d$ Set of positive semidefinite matrices.
$\mathcal{S}_d^{+,*} \subset \mathcal{S}_d^+$: Set positive definite matrices.
$\mathcal{C}^q(\mathcal{D};\mathcal{E})$: Functions defined on $\mathcal{D}$ with values in $\mathcal{E}$ with continuous derivatives up to order q.
$|\alpha| = \alpha_1 + ... + \alpha_d$ for a multi-index $\alpha = (\alpha_1, \cdots, \alpha_d) \in \mathbb{N}^d$.
$\partial_\alpha f = \partial_x^\alpha f = \partial_{x_1}^{\alpha_1} \ldots \partial_{x_d}^{\alpha_d} f(x)$, $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$.
$\|f\|_\infty = \sup_{x\in\mathcal{D}} |f(x)|$.
$\|f\|_{q,\infty} = \sum_{|\alpha|=0}^q \|\partial_\alpha f\|_\infty$.
$\|f\|_{q,1} = \sum_{0\leqslant|\alpha.|\leqslant q} \int |\partial_\alpha f(x)|.dx$.
$\mathrm{supp}(f) = \{x \in \mathcal{D}, f(x) \neq 0\}$.
$\mathcal{C}_b^q(\mathbb{R}^d) = \{f \in \mathcal{C}^q(\mathcal{D}), \|f\|_{q,\infty} < \infty\}$.
$C_c^q(\mathcal{D}) \subset C^q(\mathcal{D})$: set of functions with compact support contained in $\mathcal{D}$.
$\|f\|_\infty = \sup_x |f(x)|$.
$\int .dt$: Lebesguei Integral.
$\int .dW_t$: Ito integral.
$\int . \circ dW_t$: Stratonovich integral.

Total variation convergence

$\pi_{T,n} = \{t = kT/n, k \in \mathbb{N}\}$, $T \geqslant 0, n \in \mathbb{N}$.
$\pi_{T,n}^{\tilde{T}} = \{t \in \pi_{T,n}, t \leqslant \tilde{T}\}$, $\tilde{T}, T \geqslant 0, n \in \mathbb{N}$.
$\pi_{T,n}^{S,\tilde{T}} = \{t \in \pi_{T,n}^{\tilde{T}}, t > S\}$, $0 \leqslant S < \tilde{T}, T \geqslant 0, n \in \mathbb{N}$.
P^n : Semgroup with transition measure μ^n (see Defintion 1.2.1).
Q^n : Semgroup with transition measure ν^n (see Defintion 1.2.1).
$E_n(h,q)$: $\|(\mu^n - \nu_k^n)f\|_\infty \leqslant C\|f\|_{q,\infty}/n^{1+h}, \forall k \in \mathbb{N}^*$.
$E_n^*(h,q)$: $|\langle g, (\mu^n - \nu_k^n)f\rangle| \leqslant C\|g\|_{q,1}\|f\|_\infty/n^{1+h}, \forall k \in \mathbb{N}^*$.
$P^{n,*}, Q^{n,*}$: Adjoint semigroups in $\mathrm{L}^2(\mathbb{R}^d)$.
$R_{q,\eta}(S)$[1] : $\forall t, s \in \pi_{T,n}$, with $S \leqslant s - t$, $\|P_{t,s}^n f\|_{q,\infty} \leqslant CS^{-\eta(q)}\|f\|_\infty$. for $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ an increasing function.
$R_{q,\eta}^*(S)$[1] : $\forall t, s \in \pi_{T,n}$, with $S \leqslant s - t$, $\|P_{t,s}^{n,*} f\|_{q,1} \leqslant CS^{-\eta(q)}\|f\|_1$ for $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ an increasing function.

[1]Hypothesis satisfied by a semigroup, P^n in these examples.

$\overline{R}_{q,\eta}(S)^1 : \forall t, s \in \pi_{T,n}$,with $S \leqslant s-t$, $\sup_{|\alpha|+|\beta|\leqslant q} \|\partial_\alpha P^n_{t,s}\partial_\beta f\|_\infty \leqslant CS^{-\eta(q)}\|f\|_\infty$ for $q \in \mathbb{N}$ and $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ an increasing function.
$\|\psi\|_{1,r,\infty} = \sup_{k\in\mathbb{N}} \|\psi_k\|_{1,r,\infty} = 1 \vee \sup_{k\in\mathbb{N}} \sum_{|\alpha|=0}^{r} \sum_{|\beta|+|\gamma|=1}^{r-|\alpha|} \|\partial_x^\alpha \partial_z^\beta \partial_t^\gamma \psi_k\|_\infty$, with $(\psi_k)_{k\in\mathbb{N}} \in \mathcal{C}^\infty(\mathbb{R}^d \times \mathbb{R}^N \times \mathbb{R}_+; \mathbb{R}^d), r \in \mathbb{N}^*$.
$\mathfrak{K}_r(\psi) = (1+\|\psi\|_{1,r,\infty})\exp(\|\psi\|^2_{1,3,\infty}), r \in \mathbb{N}^*$. $M_p(Z) = 1 \vee \sup_{k\in\mathbb{N}^*} \mathbb{E}[|Z|^p]$ for a sequence of random variables $(Z_k)_{k\in\mathbb{N}}$.

MLE of the Wishart process

$X \sim WIS_d(x,\alpha,b,a)$: $X_0 = x$ and $dX_t = \left[\alpha a^\top a + bX_t + X_t b^\top\right] dt + \sqrt{X}_t dW_t a + a^\top dW_t^\top \sqrt{X}_t$ for $t > 0$.
$R_T = \int_0^T X_s ds, T \geqslant 0$.
$\overline{R}_\infty = -\alpha b^{-1}/2$.
$Q_T = \left(\int_0^T \mathrm{Tr}[X_s^{-1}]ds\right)^{-1}$.
$\overline{Q}_\infty = (\alpha - 1 - d)/(2\,\mathrm{Tr}[-b])$.
$Z_T = \log(\det[X_T]/\det[x])$.
$M_T = \int_0^T \sqrt{X_s}dW_s + \int_0^T dW_s^\top \sqrt{X}_s$.
$N_T = \int_0^T \mathrm{Tr}[(\sqrt{X_s})^{-1}dW_s]$.
$\mathcal{L}_X : \mathcal{S}_d \to \mathcal{S}_d, Y \mapsto YX + XY$, with $X \in \mathcal{S}_d$.
$\mathcal{L}_{X,a} : \mathcal{S}_d \to \mathcal{S}_d, Y \mapsto YX + XY - 2a\,\mathrm{Tr}[Y]I_d$, with $X \in \mathcal{S}_d$ and $a \in \mathbb{R}$.

Première partie

Introduction

L'objet de cette thèse est l'étude de thèmes relatifs à la modélisation de processus aléatoires. Dans une démarche classique il y a trois étapes lorsqu'on souhaite modéliser une variable par un processus stochastique : la sélection (ou la création) d'un modèle, l'étude de sa simulation et enfin le choix des paramètres à appliquer en fonction des données réelles. Dans cette thèse, nous ne traiterons pas la question du choix d'un modèle. En revanche, nous étudierons les deux autres aspects. En particulier, nous proposerons, dans la première partie, une analyse de l'erreur commise lorsqu'on utilise un schéma de discrétisation pour la simulation. En effet, ces méthodes fournissent une approximation du processus sous jacent sur une grille de temps et nécessitent donc une étude mathématique pour déterminer leur efficacité. Plus spécifiquement nous nous intéresserons à l'erreur en variation totale entre le processus et son schéma. Nous en déduirons également des propriétés sur les densités des diffusions. Dans une deuxième partie, nous aborderons la problématique du choix des paramètres. En particulier, nous étudierons l'Estimateur du Maximum de Vraisemblance (EMV) des paramètres du processus de Wishart. Cette introduction présente les motivations de ces travaux ainsi que les résultats principaux qui ont été obtenus. On s'attache ici à présenter de façon heuristique les idées conductrices qui ont guidé ces recherches. Les arguments techniques seront détaillés par la suite.

Chapitre 1

Schémas d'approximation et convergence en variation totale

L'extrême variété des domaines d'application des Équations Différentielles Stochastiques (EDS), comme la biologie, la physique ou les mathématiques financières a fortement influencé l'émergence d'un grand nombre de modèles mathématiques. Une fois le modèle choisi, l'étape suivante consiste à étudier ses propriétés. Une problématique classique repose sur le fait que, bien souvent, on ne peut étudier analytiquement les propriétés de ces modèles. On a alors fréquemment recours à des méthodes de Monte Carlo. Ces méthodes consistent à simuler un échantillon de réalisations indépendantes du processus afin notamment de calculer des espérances. Un nouvel obstacle peut alors survenir : il arrive que la simulation exacte soit trop coûteuse voire impossible à réaliser. Afin de traiter ce problème, nous utiliserons l'approximation d'EDS par des schémas de discrétisation. De cette façon, nous pourrons mettre en avant certaines propriétés de ces modèles. En particulier, pour un processus de Markov $(X_t)_{t\geqslant 0}$, nous étudierons des méthodes d'approximation de $\mathbb{E}[f(X_T)], T \geqslant 0$, pour des fonctions f mesurables et bornées. Nous obtiendrons également des estimations de la densité de X_T et de ses dérivées. Nous prouverons notamment la convergence en variation totale pour un ensemble générique de schémas. Dans ces travaux, on généralise ainsi le résultat de Bally et Talay [11], [12] qui traite le cas particulier du schéma d'Euler. De même, nous obtiendrons des estimations similaires à celles Kusuoka [44] (sous des hypothèses différentes) pour le schéma de Ninomiya Victoir.

1.1 Schémas de discrétisation pour des processus aléatoires

Dans cette Section, on considère un processus de diffusions $(X_t)_{t\geqslant 0}$ régi par une EDS. Pour $T > 0$, et $n \in \mathbb{N}^*$, nous allons nous intéresser aux schémas définis sur la grille de discrétisation homogène $\pi_{T,n} = \{t_k^n = kT/n, k \in \mathbb{N}\}$ fixé. On adopte aussi la notation $\pi_{T,n}^{\tilde{T}} = \{t \in \pi_{T,n}, t \leqslant \tilde{T}\}$ pour la version bornée de $\pi_{T,n}$ jusqu'à la date $\tilde{T}$. On précise que les résultats que nous obtiendrons restent valables pour des grilles non homogènes mais par soucis de clarté nous ne considérerons pas ce cas. Cette approche s'avère très générale puisqu'elle concerne la majorité des schémas de discrétisations présents dans la littérature. Cependant, il existe tout de même des schémas qui ne sont pas définis sur une grille fixe (voir par exemple le schéma de Kohatsu-Higa et Tankov [40] pour les processus avec sauts). Un des intérêts principaux des schémas de discrétisation réside dans leur association avec les méthodes de Monte Carlo. Ces dernières s'avèrent particulièrement efficaces pour le calcul d'espérances lorsque le calcul analytique échoue. De plus lorsque la simulation exacte est trop coûteuse, l'approximation par un schéma peut être une solution efficace. Outre l'erreur due à la méthode de Monte Carlo, dans ce type d'estimation il apparaît également une erreur due à l'approximation du processus sur la grille de temps. Dans ce travail nous nous intéresserons à cette composante. Il convient alors d'établir des procédés pour identifier la qualité de ces approximations. Une première approche consiste à considérer l'erreur forte. Cette distance mesure la différence trajectorielle entre la diffusion $(X_t)_{t\geqslant 0}$ et son schéma $(X^n_t)_{t\in\pi_{T,n}}$, et est définie par

$$\mathbb{E}[\sup_{t\in\pi^T_{T,n}} |X_t - X^n_t|]. \tag{1.1}$$

Étant donné que notre objectif est de calculer des expressions de la forme $\mathbb{E}[f(X_t)]$, pour un ensemble de fonctions test f, cette définition ne s'avère pas suffisamment pertinente. En effet, lorsqu'on souhaite estimer $\mathbb{E}[f(X_t)]$ par une méthode de Monte Carlo, pour M réalisations indépendantes $(X^{n,i}_t)_{i=1,.,M}$ de variables aléatoires de même loi que X^n_t, on a

$$\mathbb{E}[f(X_t)] = \frac{1}{M}\sum_{i=1}^{M} f(X^{n,i}_t) + \mathbb{E}[f(X_t)] - \mathbb{E}[f(X^n_t)] + \mathbb{E}[f(X^n_t)] - \frac{1}{M}\sum_{i=1}^{M} f(X^{n,i}_t). \tag{1.2}$$

Le terme $\mathbb{E}[f(X^n_t)] - \sum_{i=1}^{M} f(X^{n,i}_t)/M$ correspond à l'erreur Monte Carlo de l'estimation de $\mathbb{E}[f(X^n_t)]$. Le Théorème Central Limite implique qu'il converge vers 0 avec vitesse $1/\sqrt{M}$. Cette décomposition nous amène donc à ne pas considérer l'erreur forte mais plutôt l'erreur faible définie par

$$|\mathbb{E}[f(X_t)] - \mathbb{E}[f(X^n_t)]|, \tag{1.3}$$

où on appellera f une fonction test. On dira alors qu'un schéma $(X^n_t)_{t\in\pi_{T,n}}$ converge faiblement vers $(X_t)_{t\geqslant 0}$, pour un ensemble de fonctions test f, si pour tout $t \in \pi_{T,n}$, $|\mathbb{E}[f(X_t)] - \mathbb{E}[f(X^n_t)]| \underset{n\to\infty}{\to} 0$. Á ce stade, un sujet mérite d'être abordé. Il faut identifier la classe de fonction f la plus large possible pour laquelle on obtient la convergence faible.
Rentrons un peu plus dans les détails. On appelle ordre (faible) du schéma, l'entier h, tel que

$$|\mathbb{E}[f(X_t)] - \mathbb{E}[f(X^n_t)]| \leqslant C_f/n^h, \tag{1.4}$$

où C_f est une constante qui dépend de f. L'ordre va donc dépendre du schéma mis en œuvre. Du fait de sa pertinence pour l'étude des méthodes de Monte Carlo, l'approximation en erreur faible est un sujet déjà largement étudié dans la littérature. On peut citer par exemple les schémas d'Euler, où $h = 1$, ou de Ninomiya Victoir [57], où $h = 2$, pour des diffusions à coefficients réguliers et des schémas construits à partir de variables aléatoires Gaussiennes. Les preuves de ces résultats (voir [55], [68] ou encore [39] pour le schéma d'Euler) reposent sur des développements pertinents de l'erreur faible en temps court, c'est à dire entre deux points de la grille de discrétisation, combinés avec les Équations Différentielles Partielles (EDP) induites par le Théorème de Feynman Kac. On peut aussi se reporter au travail d'Alfonsi [3] pour des schémas sur le processus de Cox Ingersoll Ross (CIR) inspirés de [57]. Outre le fait qu'il ne traite pas une diffusion à coefficients réguliers, un point pertinent dans son approche repose sur le fait qu'il ne considère plus nécessairement des variables aléatoires Gaussiennes dans la construction de son schéma mais une classe de variables aléatoires qui vérifient des conditions de moments. Cette variabilité dans les possibilités choix des variables aléatoires qui peuvent être utilisées dans la construction de schémas de discrétisation avait déjà été mis en avant dans les travaux initiaux de Talay [67].

Cependant, dans tous ces travaux, la constante C_f dépend non seulement de f mais aussi de ses dérivées. Un autre objectif de l'étude des schémas de discrétisation repose donc sur la possibilité d'élargir cet espace de fonctions test tout en conservant l'ordre du schéma. Notamment on veut montrer que (1.5) reste valable pour $C_f = C\|f\|_\infty$. À ce jour, certains résultats de cette forme ont déjà été obtenus mais ils concernent toujours un schéma particulier. Pour le schéma d'Euler, la preuve revient à Bally et Talay [11], [12] tandis que pour le schéma de Ninomiya Victoir on peut se reporter à Kusuoka [44]. Les démonstrations de ces résultats exploitent le caractère Gaussien des variables aléatoires mises en œuvre dans les schémas et profitent ainsi du Calcul de Malliavin. Ici, on ne fera pas une telle hypothèse mais on considérera une classe générale de variables aléatoires. La calcul de Malliavin ne s'appliquant pas dans ce cas, on utilisera une de ses variantes introduite dans [9]. De même, on donnera des conditions suffisantes sur le schéma pour obtenir la convergence sans pour autant spécifier ce schéma. La démarche considérée dans cette thèse vise ainsi à étendre les résultats de [11], [12] et [44] à un ensemble plus général de schémas incluant ces exemples. On obtiendra donc, pour toute fonction f mesurable et bornée,

$$|\mathbb{E}[f(X_t)] - \mathbb{E}[f(X_t^n)]| \leqslant C\|f\|_\infty / n^h, \tag{1.5}$$

pour tout schéma $(X_t^n)_{t\in\pi_{T,n}}$ de cette classe. On dira alors que le schéma converge en variation totale vers la diffusion. Afin d'obtenir ces résultats, nous emploierons l'approche par semigroupes qui correspond à une autre vision de l'approche par EDPs relative au Théorème de Feynman Kac.

1.2 Présentation du problème

On rappelle que $T > 0$, $n \in \mathbb{N}^*$ et que les schémas d'approximation que nous étudions sont définis sur la grille de temps $\pi_{T,n}$. Considérons une chaine de Markov d-dimensionnelle, $(X_t^n)_{t\in\pi_{T,n}}$ de la forme

$$X^n_{t^n_{k+1}} = \psi_k(X^n_{t^n_k}, \frac{Z_{k+1}}{\sqrt{n}}, \delta^n_{k+1}), \qquad t^n_k = kT/n, \quad k \in \mathbb{N}, \tag{1.6}$$

où $\psi_k : \mathbb{R}^d \times \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}^d$ vérifie $\psi_k(x,0,0) = x$, $Z_{k+1} \in \mathbb{R}^N,\ k \in \mathbb{N}$, est une suite de variables aléatoires centrées et $\sup_{k\in\mathbb{N}^*} \delta^n_k \leqslant C/n$.
Par exemple, si on s'intéresse au schéma d'Euler de la diffusion à coefficients réguliers,

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t,$$

on a $\psi_k(x,z,t) = x + b(x)t + \sigma(x)z$ où $Z_k \sim \sqrt{T}\mathcal{N}(0,1)$ avec $\mathcal{N}(0,1)$ la loi Normale centrée réduite et $\delta^n_{k+1} = t^n_{k+1} - t^n_k = T/n$.
On notera $(Q^n_t)_{t\in\pi_{T,n}}$ le semigroupe associé à la chaîne de Markov $(X^n_t)_{t\in\pi_{T,n}}$ et $\nu^n_{k+1}(x,dy) = \mathbb{P}(X^n_{t^n_{k+1}} \in dy | X^n_{t^n_k} = x)$, sa suite de probabilités de transition entre t^n_k et t^n_{k+1}. On rappelle que pour $t \in \pi_{T,n}$, le semigroupe de la chaine de Markov est défini par $Q^n_t f(x) = \mathbb{E}[f(X^n_t)|X^n_0 = x]$. De plus pour $t, s \in \pi_{T,n}, t \leqslant s$, on introduit le semigroupe généralisé $Q^n_{t,s}f(x) = \mathbb{E}[f(X^n_s)|X^n_t = x]$. De façon similaire, on considère le processus de Markov en temps continu $(X_t)_{t\geqslant 0}$. On appellera $(P_t)_{t\geqslant 0}$ son semigroupe et $\mu^n_{k+1}(x,dy) = \mathbb{P}(X_{t^n_{k+1}} \in dy | X_{t^n_k} = x)$, sa suite de probabilités de transition entre t^n_k et t^n_{k+1}. L'approximation en erreur faible est un thème largement exploré dans la littérature et nous initierons notre réflexion avec le résultat suivant. On suppose qu'il existe $h > 0,\ q \in \mathbb{N}$ tels que pour tout $f \in \mathcal{C}^q(\mathbb{R}^d)$, $k \in \mathbb{N}^*$ et $x \in \mathbb{R}^d$,

$$|\mu^n_k f(x) - \nu^n_k f(x)| = |\textstyle\int f(y)\mu^n_k(x,dy) - \int f(y)\nu^n_k(x,dy)| \leqslant C\|f\|_{q,\infty}/n^{h+1} \tag{1.7}$$

où $\|f\|_{q,\infty}$ est la norme infinie de la fonction f et de ses dérivées jusqu'à l'ordre q. Alors, pour tout $T > 0$, il existe $C \geqslant 1$ tel que

$$\sup_{t\in\pi^T_{T,n}} \|P_t f - Q^n_t f\|_\infty \leqslant C\|f\|_{q,\infty}/n^h. \tag{1.8}$$

Dans ce cas, on dira que $(X^n_t)_{t\in\pi_{T,n}}$ est un schéma d'approximation d'ordre faible h pour le processus de Markov $(X_t)_{t\geqslant 0}$ pour l'ensemble des fonctions test $f\in\mathcal{C}^q_b(\mathbb{R}^d;\mathbb{R})$. La valeur de h mesure donc la qualité de l'approximation tandis que celle de q représente la régularité nécessaire sur la fonction test f afin que le schéma soit efficace.

Dans le cas du schéma d'Euler pour les diffusions, ce résultat a été initialement prouvé avec $h=1$ et $q=4$ dans les articles de Milstein [55] et de Talay and Tubaro [68] (voir aussi [39]). D'autres auteurs se sont ensuite attachés à prouver ce résultat dans différentes situations : les diffusions avec sauts (voir [63], [36]) ou celles avec des conditions aux bornes (voir [30], [18], [31]). Le lecteur intéressé pourra trouver une revue complète du sujet dans [38]. Par la suite, des schémas d'ordre faible plus élevé ont été mis en œuvre. En utilisant des méthodes de cubature, des schémas vérifiant (1.8) avec $h=2$ ont été développés par Kusuoka [43], Lyons [53], Ninomiya et Victoir [57], Alfonsi [3], Kohatsu-Higa et Tankov [40]. On peut aussi trouver un schéma d'ordre $h=3$ dans [3] qui repose sur une approche similaire. Cependant dans tout ces travaux, l'ordre q nécessaire à la convergence faible est supérieur à un.

Un autre résultat concerne alors la régularité nécessaire sur la fonction test f. L'objectif est d'obtenir (1.8) avec $\|f\|_{q,\infty}$ remplacé par $\|f\|_\infty$, pour toute fonction f mesurable et bornée. On dira alors que $(X^n_t)_{t\in\pi_{T,n}}$ converge en variation totale vers le processus de Markov $(X_t)_{t\geqslant 0}$. Dans le cas du schéma d'Euler, Bally et Talay [11], [12] sont les premiers à avoir prouvé ce résultat en utilisant les formules d'intégration par parties du calcul de Malliavin. Plus tard, par des techniques d'intégration par parties classique, Guyon [34] a étendu ce résultat au cas où f est une distribution tempérée. Ensuite, Konakov, Menozzi and Molchanov [41], [42] ont démontré des théorèmes de limite locale pour des approximations de chaines de Markov en utilisant une méthode paramétrique. Enfin, plus récemment, Kusuoka [44] a obtenu une estimation de l'erreur en variation totale pour le schéma de Ninomiya Victoir (h=2) sous une condition de type Hörmander.

Nous montrerons que, si l'hypothèse (1.7) est vérifiée, nous obtenons (1.8) avec $\|f\|_\infty$ sous une condition d'ellipticité. De plus, nous n'utiliserons pas de représentation Gaussienne pour la suite $(Z_k)_{k\in\mathbb{N}^*}$ et ainsi le calcul de Malliavin classique ne sera pas applicable. Pour démontrer ce résultat, nous emploierons un calcul de Malliavin abstrait initialement introduit dans [9]. La convergence en variation totale va donc s'appliquer à un ensemble de schémas et l'estimation de l'erreur faible pourra être considérée comme un principe d'invariance.

1.3 Propriétés de régularisation et de convergences abstraites

Afin d'obtenir (1.8), il est nécessaire d'avoir de fortes propriétés de régularité sur la fonction test f. Nous allons montrer ici que sous des hypothèses de régularisation des semigroupes concernés, il est possible de s'affranchir de ces conditions et de remplacer $\|f\|_{q,\infty}$ par $\|f\|_\infty$. Nous emploierons ensuite ces résultats pour traiter les cas où les fonctions ψ_k sont régulières. Nous utiliserons la notation

$$|\mu^n_k| := \sup_{x\in\mathbb{R}^d}\sup_{\|f\|_\infty\leqslant 1}\Big|\int_{\mathbb{R}^d} f(y)\mu^n_k(x,dy)\Big|. \tag{1.9}$$

Tout d'abord on suppose que

$$\sup_{k\in\mathbb{N}^*}|\mu^n_k| + \sup_{k\in\mathbb{N}^*}|\nu^n_k| < \infty. \tag{1.10}$$

La preuve de la convergence en variation totale repose alors sur deux éléments. Le premier, qui est aussi nécessaire pour obtenir (1.8), est l'approximation en temps court. Ici nous aurons besoin d'une hypothèse d'approximation en temps court "directe" : soit $q \in \mathbb{N}$. Il existe une constante $C > 0$ (dépendant de q seulement) telle que, pour tout $k \in \mathbb{N}^*$,

$$E_n(h, q) \qquad \|(\mu_k^n - \nu_k^n)f\|_\infty \leqslant C\|f\|_{q,\infty}/n^{h+1}. \tag{1.11}$$

Une hypothèse d'approximation en temps court "adjointe" sera également nécessaire pour atteindre notre résultat : soit $q \in \mathbb{N}$. Il existe une constante $C \geqslant 1$ telle que, pour toute fonction f mesurable bornée et $g \in \mathcal{C}^q(\mathbb{R}^d)$, on a

$$E_n^*(h, q) \qquad |\langle g, (\mu_k^n - \nu_k^n)f\rangle| \leqslant C\|g\|_{q,1}\|f\|_\infty/n^{h+1}, \tag{1.12}$$

où $\langle g, f\rangle = \int_{\mathbb{R}^d} g(x)f(x)dx$ est le produit scalaire usuel dans $\mathrm{L}^2(\mathbb{R}^d)$.
Une fois ces hypothèses d'approximation en temps court établies, nous pouvons aborder les hypothèses de régularisation des semigroupes. On commence par introduire l'hypothèse de régularisation "classique" : soient $q \in \mathbb{N}$, $S > 0$ et $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ une fonction croissante. Il existe une constante $C \geqslant 1$ telle que

$$R_{q,\eta}(S) \qquad \forall t, s \in \pi_{T,n}, \text{ avec } S \leqslant s - t, \quad \|P_{t,s}^n f\|_{q,\infty} \leqslant \frac{C}{S^{\eta(q)}}\|f\|_\infty. \tag{1.13}$$

Ensuite, on considère une hypothèse de régularisation "adjointe". On suppose l'existence d'un semigroupe adjoint $P_{t,s}^{n,*}$ qui vérifie

$$\left\langle P_{t,s}^{n,*} g, f \right\rangle = \left\langle g, P_{t,s}^n f \right\rangle$$

pour toute fonction f mesurable bornée et toute fonction $g \in \mathcal{C}_c^\infty(\mathbb{R}^d)$. On suppose alors que $P_{t,s}^{n,*}$ satisfait

$$R_{q,\eta}^*(S) \qquad \forall t, s \in \pi_{T,n}, \text{ avec } S \leqslant s - t, \quad \|P_{t,s}^{n,*} f\|_{q,1} \leqslant \frac{C}{S^{\eta(q)}}\|f\|_1. \tag{1.14}$$

Dans la pratique, il n'est pas toujours aisé de mettre en évidence cette propriété. Nous proposons donc une condition suffisante à l'obtention de $R_{q,\eta}^*(S)$: pour tout multi-indice $|\alpha| \leqslant q$, on a

$$\forall t, s \in \pi_{T,n}, \text{ avec } S \leqslant s - t, \quad \|P_{t,s}^n \partial_\alpha f\|_\infty \leqslant \frac{C}{S^{\eta(q)}}\|f\|_\infty. \tag{1.15}$$

En effet, par un calcul direct,

$$\begin{aligned}
\|\partial_\alpha P_{t,s}^{n,*} f\|_1 &\leqslant \sup_{\|g\|_\infty \leqslant 1} |\left\langle \partial_\alpha P_{t,s}^{n,*} f, g \right\rangle| = \sup_{\|g\|_\infty \leqslant 1} |\left\langle f, P_{t,s}^n(\partial_\alpha g) \right\rangle| \\
&\leqslant \|f\|_1 \sup_{\|g\|_\infty \leqslant 1} \|P_{t,s}^n(\partial_\alpha g)\|_\infty \leqslant \frac{C}{S^{\eta(q)}}\|f\|_1.
\end{aligned}$$

Finalement, pour $\tilde{T} \geqslant 0$ et $S \in [0, \tilde{T})$ on notera $\pi_{T,n}^{S,\tilde{T}} = \{t \in \pi_{T,n}^{\tilde{T}}, t > S\}$. Une fois toutes ces hypothèses à notre disposition, nous allons pouvoir fournir un premier résultat, que nous appellerons "abstrait", de convergence en variation totale entre deux semigroupes P^n et Q^n.

Proposition 1.3.1. *On fixe $q \in \mathbb{N}$, $h \geqslant 0$, $S \in [T/n, T/2)$ et $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ une fonction croissante.. On suppose que les hypothèses $E_n(h, q)$ (voir (1.11)) et $E_n^*(h, q)$ (voir (1.12)) sont satisfaites entre P^n et Q^n. On suppose aussi que P^n vérifie $R_{q,\eta}(S)$ (voir (1.13)) et Q^n satisfait $R_{q,\eta}^*(S)$ (voir (1.14)). Alors,*

$$\sup_{t \in \pi_{T,n}^{2S,T}} \|P_t^n f - Q_t^n f\|_\infty \leqslant \frac{C}{S^{\eta(q)}}\|f\|_\infty/n^h. \tag{1.16}$$

Désormais, nous allons expliciter des conditions nécessaires pour obtenir les propriétés de régularisation dans le cas où X^n, la chaîne de Markov associée au semigroupe Q^n, a la forme (1.6).

1.4 Formules d'intégration par parties - L'objet de base

Après avoir énoncé ce résultat abstrait, l'objectif suivant consiste à donner des conditions suffisantes pour obtenir $R_{q,\eta}$, $R^*_{q,\eta}$ et $\overline{R}_{q,\eta}$. Afin de démontrer ces propriétés, nous emploierons des formules d'intégration par parties basées sur les bruits $Z_k \in \mathbb{R}^N$. Ces formules sont semblables à celles que l'on trouve dans le calcul de Malliavin. La différence ici, c'est qu'on ne suppose pas que les variables aléatoires Z_k ont une distribution Gaussienne et ainsi le calcul de Malliavin standard n'est plus adapté. Pour cela, nous allons supposer que chaque Z_k est localement bornée inférieurement par la mesure de Lebesgue, c'est à dire : il existe une suite $z_{*,k} \in \mathbb{R}^N$, $k \in \mathbb{N}^*$, r_* et $\varepsilon_* > 0$ tels que pour tout ensemble mesurable $A \subset B_{r_*}(z_{*,k})$, on a

$$\mathbb{P}(Z_k \in A) \geqslant \varepsilon_* \lambda(A) \tag{1.17}$$

où λ est la mesure de Lebesgue. Dans le cas où cette propriété est vérifiée il est possible d'appliquer une méthode de "splitting" afin de représenter Z_k de la façon suivante :

$$\frac{Z_k}{\sqrt{n}} = \chi_k U_k + (1-\chi_k) V_k,$$

où χ_k, U_k, V_k sont des variables aléatoires indépendantes, χ_k est une variable aléatoire de Bernoulli et $\sqrt{n}U_k \sim \varphi_{r_*}(u)du$ avec $\varphi_{r_*} \in \mathcal{C}^\infty(\mathbb{R}^N)$. Dans ce cas, nous pourrons utiliser un calcul de Malliavin abstrait basé sur la densité de U_k, et développé dans [9] et [6]. Ainsi, nous obtiendrons des formules d'intégration par parties qui vont nous permettre de prouver (1.16). La clé de notre approche repose sur le fait que la densité φ_{r_*} de $\sqrt{n}U_k$ est régulière et qu'on contrôle ses dérivées logarithmiques. En exploitant ces propriétés, il va être possible de construire des formules d'intégration par parties et d'obtenir des bornes pertinentes pour les normes des poids qui vont apparaître des ces formules. De telles variantes du calcul de Malliavin ont déjà été utilisées par le passé. Une méthode de splitting similaire apparaît dans Nourdin and Poly [59] (voir aussi [58] et [48]). Ils emploient ce qu'on appelle le "Γ calculus" introduit par Bakry, Gentil et Ledoux [5]. En s'exprimant de façon heuristique, on peut distinguer notre approche de celle de [5] de la façon suivante : notre démarche est semblable à celle des "fonctionnelles simples" dans le calcul de Malliavin et utilise l'opérateur de dérivation comme objet central. En revanche, l'objet de base du "Γ calculus" est l'opérateur d'Ornstein Uhlenbeck.

1.5 Convergence en variation totale

Afin d'être en mesure d'énoncer notre résultat principal, nous introduisons des hypothèses additionnelles.

$$\forall p \in \mathbb{N}, \quad \sup_{k\in\mathbb{N}} \mathbb{E}[|Z_k|^p] < \infty, \tag{1.18}$$

$$\forall r \in \mathbb{N}^*, \quad \sup_{k\in\mathbb{N}} \|\psi_k\|_{1,r,\infty} = \sum_{|\alpha|=0}^{r} \sum_{|\beta|+|\gamma|=1}^{r-|\alpha|} \|\partial_x^\alpha \partial_z^\beta \partial_t^\gamma \psi_k\|_\infty < \infty, \tag{1.19}$$

$$\exists \lambda_* > 0,\ \forall k \in \mathbb{N}, \quad \inf_{x\in\mathbb{R}^d} \inf_{|\eta|=1} \sum_{i=1}^{N} \langle \partial_{z_i}\psi_k(x,0,0), \eta\rangle^2 \geqslant \lambda_*. \tag{1.20}$$

Et pour $r \in \mathbb{N}^*$, on note

$$\mathfrak{K}_r(\psi) = (1 + \|\psi\|_{1,r,\infty}) \exp(\|\psi\|_{1,3,\infty}^2). \tag{1.21}$$

De plus, pour obtenir une estimation des dérivées de tout ordre de la densité de X_t, $t > 0$, nous introduisons une modification de $(X_t^n)_{t \in \pi_{T,n}}$:

$$\forall t \in \pi_{T,n}, \quad X_t^{n,\theta}(x) = \frac{1}{n^\theta} G + X_t^n(x), \tag{1.22}$$

où G est une variable aléatoire normale centrée réduite indépendante de $X_{t_k^n}^n$ et $\theta \geqslant h + 1$. On utilise ici la notation $X_t^n(x)$ pour la chaine de Markov partant de x : $X_0^n(x) = x$. On introduit alors

$$Q_t^{n,\theta}(x, dy) = \mathbb{P}(X_t^{n,\theta}(x) \in dy) = p_t^{n,\theta}(x, y) dy. \tag{1.23}$$

Théorème 1.5.1. *On fixe $q \in \mathbb{N}$ et $h > 0$. Pour un $n \in \mathbb{N}^*$ donné, on considère les semigroupes Markovien $(P_t)_{t \geqslant 0}$ et $(Q_t^n)_{t \in \pi_{T,n}}$, définie plus haut. Alors, il existe $n_0 \in \mathbb{N}^*$ tel que pour tout $n \geqslant n_0$, $S \in \pi_{T,n}$, $S \in [T/n, T/2)$, les propriétés suivantes sont vérifiées.*

A. *On suppose que les hypothèses (1.17), (1.18), (1.19) et (1.20) sont vérifiées. De plus on suppose que $E_m(h, q)$ (voir (1.11)) et $E_m^*(h, q)$ (voir (1.12)) sont satisfaites entre $(P_t^m)_{t \in \pi_T^m} = (P_t)_{t \in \pi_T^m}$ et $(Q_t^m)_{t \in \pi_T^m}$ pour tout $m \geqslant n$. Alors, il existe $l \in \mathbb{N}^*$, $C \geqslant 1$, qui dépendent de q, T et des moments de Z, tels que*

$$\sup_{t \in \pi_{T,n}^{2S,T}} \|P_t f - Q_t^n f\|_\infty \leqslant C \frac{\mathfrak{K}_{q+3}(\psi)^l}{(\lambda_* S)^{\eta(q)}} \|f\|_\infty / n^h. \tag{1.24}$$

avec $\eta(q) = q(q+1)$.

B. *De plus, pour tout $t > 0, P_t(x, dy) = p_t(x, y) dy$ avec $(x, y) \to p_t(x, y)$ qui appartient à $\mathcal{C}^\infty(\mathbb{R}^d \times \mathbb{R}^d)$.*

C. *Enfin, soit $\theta \geqslant h + 1$. Alors, on a $Q_t^{n,\theta}(x, dy) = p_t^{n,\theta}(x, y) dy$ et il existe $l \in \mathbb{N}^*$ tel que pour tout $R, \varepsilon > 0$, $x_0, y_0 \in \mathbb{R}^d$ et tout multi-indice α, β avec $|\alpha| + |\beta| = u$,*

$$\sup_{t \in \pi_{T,n}^{2S,T}} \sup_{(x,y) \in \overline{B}_R(x_0,y_0)} |\partial_x^\alpha \partial_y^\beta p_t(x, y) - \partial_x^\alpha \partial_y^\beta p_t^{n,\theta}(x, y)| \leqslant C \frac{\mathfrak{K}_{q+3}(\psi)^l}{(\lambda_* S)^{\eta(p_{u,\varepsilon} \vee q)}} / n^{h(1-\varepsilon)}, \tag{1.25}$$

où C est une constante qui dépend de R, x_0, y_0, T, $|\alpha| + |\beta|$ et des moments Z et $p_{u,\varepsilon} = (u + 2d + 1 + 2\lceil (1-\varepsilon)(u+d)/(2\varepsilon) \rceil)$. De plus on utilise la notation $\overline{B}_R(x_0, y_0) = \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d, |(x, y) - (x_0, y_0)| \leqslant R\}$.

On remarque que (1.24) signifie la convergence en variation totale entre $(P_t)_{t \geqslant 0}$ et $(Q_t^n)_{t \in \pi_{T,n}}$. La preuve de (1.24) s'avère élémentaire une fois qu'on a prouvé les propriétés de régularisation à l'aide du calcul de Malliavin abstrait introduit dans [9]. A contrario, l'estimation (1.25) repose sur un résultat non trivial d'interpolation établi dans [8]. Il faut aussi remarquer que l'estimation (1.25) est sous-optimale (car $\varepsilon > 0$). Enfin, un intérêt remarquable de ce résultat repose sur le fait qu'il n'y a aucune hypothèse de régularité à vérifier sur le semigroupe $(P_t^n)_{t \in \pi_{T,n}} = (P_t)_{t \in \pi_{T,n}}$ contrairement à la Proposition 1.3.1. Ici la seule hypothèse qu'il doit satisfaire est (1.10) ce qui s'avère relativement utile dans les applications concrètes.

De plus, $(Z_k)_{k\in\mathbb{N}^*}$ est une suite de variables aléatoires de loi générale (executé pour les hypothèses (1.17) et (1.18)), et ainsi le Théorème 1.5.1 peut être vue comme un principe d'invariance. On peut alors se demander si l'hypothèse (1.17) n'est pas trop restrictive. Une analogie avec le Théorème de Prokhorov établi dans [62] peut nous convaincre que ce n'est pas le cas. Dans ce travail, il étudie la convergence du Théorème Central Limite (TCL), non pas en loi mais en variation totale et fournit une condition nécessaire et suffisante pour obtenir cette convergence. Nous donnons l'énoncé de son résultat.

Théorème. *Prokhorov (1952).*
Pour une suite de variables aléatoires indépendantes et identiquement distribuées (centrées et réduites) $(Y_k)_{k\in\mathbb{N}}$, alors $(Y_1+.+Y_n)/\sqrt{n}$ converge en variation totale vers la distribution Gaussienne standard si et seulement si il existe $m\in\mathbb{N}$ tel que $Y_1+.+Y_m$ a une composante absolument continue par rapport à la mesure de Lebesgue (c'est à dire $\mathbb{P}(Y_1+\ldots+Y_m\in dx)=\gamma(dx)+g(x)dx$ où γ est une mesure positive et g une fonction mesurable positive).

Nous allons montrer qu'on peut retrouver la condition suffisante pour la convergence dans ce Théorème par une application du Théorème 1.5.1. Tout d'abord, on précise qu'il est possible de prouver que la somme de deux variables aléatoires $\tilde{Y}_1$ et $\tilde{Y}_2$ ayant un composante absolument continue par rapport à la mesure de Lebesgue possède une composante continue par rapport à la mesure de Lebesgue, c'est à dire : $\mathbb{P}(\tilde{Y}=\tilde{Y}_1+\tilde{Y}_2\in dx)=\gamma(dx)+g(x)dx$ où γ est une mesure positive et g une fonction continue positive. Cette propriété va s'avérer cruciale pour notre application puisqu'elle nous permet de nous placer dans le cadre dans lequel nous avons établi le Théorème 1.5.1. En effet, on peut montrer aisément que $\mathbb{P}(\tilde{Y}\in dx)=\gamma(dx)+g(x)dx$ si et seulement si $\tilde{Y}$ vérifie (1.17).
Revenons maintenant à la preuve de la convergence en variation totale. On suppose qu'il existe $m\in\mathbb{N}$ tel que $Y_1+.+Y_m$ a une composante absolument continue par rapport à la mesure de Lebesgue. Dans ce cas $\tilde{Z}_k=Y_{2km+1}+\ldots+Y_{2km+2m}$ a une composante continu par rapport à la mesure de Lebesgue et satisfait ainsi (1.17). Maintenant on choisit $\psi_k(x,z,t)=x+z$ et $Z_k=\tilde{Z}_k$, puis on applique le Théorème 1.5.1. On retrouve alors la convergence en variation totale de $(Y_1+.+Y_n)/\sqrt{n}$ vers la Gaussienne centrée réduite.
Notre approche fournit donc une preuve alternative de la condition suffisante du Théorème de Prokhorov. Comme il s'agit également d'une condition nécessaire dans [62], cette analogie, dans le cas du TCL, nous amène a penser que la condition (1.17) se rapproche d'une condition nécessaire pour obtenir la convergence en variation totale dans le Théorème 1.5.1.

Dans ces travaux, on illustrera le Théorème 1.5.1 en construisant des schémas inspirés de l'approche de Ninomiya Victoir et qui convergent avec ordres $h=2$ et $h=3$.

Le schéma de Ninomiya Victoir.
Nous allons appliquer le Théorème 1.5.1 dans le cas ou X^n est le schéma de Ninomiya Victoir d'un processus de diffusion à coefficients réguliers. Il s'agit d'une variante du résultat obtenu par Kusuoka [44] qui suppose que les Z_k sont Gaussiennes et bénéficie ainsi des techniques de calcul de Malliavin classiques. De plus il considère une hypothèse de type Hörmander alors que nos résultats sont valables sous une hypothèse d'ellipticité. Étant donné que dans notre approche Z_k peut avoir une distribution arbitraire, comme il l'a déjà été mentionné précédemment, notre résultat peut être interprété comme un principe d'invariance. On s'intéresse au processus d-dimensionnel

$$dX_t=\sum_{i=1}^{N}V_i(X_t)\circ dW_t^i+V_0(X_t)dt \tag{1.26}$$

où $V_0, V_i \in \mathcal{C}_b^\infty(\mathbb{R}^d;\mathbb{R}^d),\ i=1,\ldots,N$ et $W=(W^1,\ldots,W^N)$ est un mouvement Brownien standard et $\circ dW_t^i$ représente l'intégrale de Stratonovich par rapport à W^i. Le générateur infinitésimal d'un tel processus est donné par

$$A = V_0 + \frac{1}{2}\sum_{k=1}^N V_k^2, \tag{1.27}$$

avec la notation $Vf(x) = \langle V(x), \nabla f(x)\rangle$. On définit maintenant $\exp(V)(x) := \Phi_V(x,1)$ où Φ_V est la solution de l'Equation Différentielle Ordinaire (EDO) $\Phi_V(x,t) = x + \int_0^t V(\Phi_V(x,s))ds$. Nous sommes désormais en mesure de présenter le schéma de Ninomiya Victoir. Soit $\rho_k,\ k\in\mathbb{N}$ une suite de variables aléatoires indépendantes suivant une loi de Bernoulli. On définit $\psi_k : \mathbb{R}^d \times \mathbb{R}^{N+1} \to \mathbb{R}^d$ de la façon suivante

$$\begin{aligned}
\psi_k(x,w^1,w^0) &= \exp(w^0V_0)\circ\exp(w^{1,1}V_1)\circ\cdot\circ\exp(w^{1,N}V_N)\circ\exp(w^0V_0)(x), \quad \text{si } \rho_k = 1, \quad (1.28)\\
\psi_k(x,w^1,w^0) &= \exp(w^0V_0)\circ\exp(w^{1,N}V_N)\circ\cdot\circ\exp(w^{1,1}V_1)\circ\exp(w^0V_0)(x), \quad \text{si } \rho_k = -1.
\end{aligned}$$

On appliquera ce schéma avec $w_{k+1}^0 = T/2n$ et $w_{k+1}^1 = (w_{k+1}^{1,i})_{i=1,\cdot,N}$ où $w_{k+1}^{1,i} = Z_{k+1}^i\sqrt{T/n}$. De plus, on supposera que les $Z_k^i,\ i=1,\ldots,N,\ k\in\mathbb{N}^*$ sont des variables aléatoires indépendantes qui vérifient (1.17) et (1.18) ainsi que la condition suivante sur les moments :

$$\mathbb{E}[Z_k^i] = \mathbb{E}[(Z_k^i)^3] = \mathbb{E}[(Z_k^i)^5] = 0, \qquad \mathbb{E}[(Z_k^i)^2] = 1, \qquad \mathbb{E}[(Z_k^i)^4] = 6. \tag{1.29}$$

Dans l'article original de Ninomiya Victoir, on suppose que les Z_k^i sont des variables aléatoires centrées et Gaussiennes et donc vérifient (1.17). La nouveauté dans cette application de notre théorème repose sur le fait qu'il n'est pas nécessaire d'avoir une loi particulière pour les Z_k mais seulement les hypothèses plus faibles (1.17) et (1.29). On rappelle que $t_k^n = kT/n$. Une étape de notre schéma (entre les temps t_k^n et t_{k+1}^n) est alors donnée par

$$X_{t_{k+1}^n}^n = \psi_{k+1}(X_{t_k^n}^n, w_{k+1}^1, w_{k+1}^0). \tag{1.30}$$

Sous une condition d'ellipticité, nous allons pouvoir fournir une estimation de la distance en variation totale entre le processus de diffusion de la forme (1.26) et son schéma de Ninomiya Victoir (1.30).

Théorème 1.5.2. *On rappelle que $V_i \in \mathcal{C}_b^\infty(\mathbb{R}^d;\mathbb{R}^d), i=0,\ldots,N$ et on suppose que*

$$\inf_{|\xi|=1}\sum_{i=1}^N \langle V_i(x),\xi\rangle^2 \geqslant \lambda_* > 0 \qquad \forall x\in\mathbb{R}^d. \tag{1.31}$$

Soit $S\in(0,T/2)$. Alors, il existe $n_0\in\mathbb{N}^*$ tel que pour tout $n\geqslant n_0$, il existe $l\in\mathbb{N}^*$, $C\geqslant 1$ tels que pour toute fonction mesurable et bornée $f:\mathbb{R}^d\to\mathbb{R}$, on a

$$\sup_{t\in\pi_{T,n}^{2S,T}} |\mathbb{E}[f(X_t)] - \mathbb{E}[f(X_t^n)]| \leqslant C\frac{C_6(V)^l\mathfrak{K}_9(\psi)^l}{(\lambda_*S)^{\eta(6)}}\|f\|_\infty/n^2. \tag{1.32}$$

avec $C_q(V) := \sup_{i=0,.,N}\|V_i\|_{q,\infty}$ et pour $q\in\mathbb{N}, r\in\mathbb{N}^*$, $\eta(q) = q(q+1)$ et $\mathfrak{K}_r(\psi)$ est défini par (1.21).

Sous l'hypothèse $V_i\in\mathcal{C}_b^\infty(\mathbb{R}^d;\mathbb{R}^d), i=0,\ldots,N$, on peut montrer que $\Gamma_{10}(\psi)<\infty$. Ainsi (1.32) signifie la convergence en variation totale du schéma de Ninomiya Victoir avec ordre 2. A la vue du Théorème 1.5.1 on peut alors noter deux éléments.

Tout d'abord, pour tout $t \geqslant 0$, on a l'existence d'une densité pour X_t. De plus il est possible de construire un schéma $X_t^{n,\theta}$ (voir (1.22)) tel que les dérivées de tout ordre de la densité de $X_t^{n,\theta}$ convergent vers celles de la densité de X_t sur tout compact de $\mathbb{R}^d \times \mathbb{R}^d$.

Ensuite, on remarque que le terme de domination dans (1.32) dépend en partie de $C_6(V)$. Cette constante apparaît lors des approximations en temps court (1.11) et (1.12). L'autre partie de ce terme apparaît déjà dans le Théorème 1.5.1 et est dûe à la régularisation des semigroupes (avec $q = 6$ dans (1.24)). Par une démarche similaire on s'attend donc à pouvoir prouver la convergence en variation totale pour des schémas construits de façon analogue à (1.28) et d'ordre faible encore plus élevé ($h > 2$). La difficulté principale reposera alors dans l'approximation en temps court et non dans la régularisation des semigroupes.
Nous illustrons cette remarque dans le Chapitre 2 de la Partie II. Nous nous intéresserons à la convergence d'un schéma d'ordre $h = 3$ introduit dans [3] dans le cas particulier $d = N = 1$. Dans cette approche, l'existence et la convergence du schéma pour des fonctions test régulières sont problématiques. En effet, la composition (1.28) est liée à la solution d'un problème algébrique visant à éliminer les termes d'ordres inférieurs (ou égaux) à 2 dans le développement en temps court de l'erreur faible. Ainsi, pour prouver la convergence à l'ordre $h = 3$, il faut éliminer les termes d'ordres inférieurs à 3 dans ce développement. Or, il s'avère que cette opération est bien plus complexe que dans le cas $h = 2$ et sa réussite dépend de propriétés de commutation des opérateurs $V_i, i = 1, \ldots, N$. En se plaçant dans le cas $d = N = 1$, nous allons fournir une condition nécessaire et suffisante explicite pour obtenir $E_n(3, 8)$ et $E_n^*(3, 8)$. Sans perte de généralité, on suppose que V_1 est une fonction à valeur dans $\mathbb{R}_+^*$. Ainsi, on montre que sous des hypothèses similaires à celles du Théorème 1.5.2 et si de plus la fonction $x \mapsto V_0(x)/V_1(x)$ est croissante, alors il est possible de construire un schéma X_t^n de façon analogue à (1.28) et qui converge en variation totale vers X avec ordre 3 en n. De plus, on a

$$\sup_{t \in \pi_{T,n}^{2S,T}} |\mathbb{E}[f(X_t)] - \mathbb{E}[f(X_t^n)]| \leqslant C \frac{C_8(V)^l \mathfrak{K}_{11}(\psi)^l}{(\lambda_* S)^{\eta(8)}} \|f\|_\infty / n^3, \tag{1.33}$$

où $C_k(V) := \sup_{i=0,1} \|V_i\|_{k,\infty} + \|\tilde{V}\|_{k,\infty}$, avec

$$\tilde{V}(x) = \sqrt{|V_1(x)(V_1(x)\partial_x V_0(x) - \partial_x V_1(x) V_0(x))|}, \quad x \in \mathbb{R},$$

est la constante qui apparaît dans $E_n(3, 8)$ et $E_n^*(3, 8)$. Ce résultat est donné dans le Théorème 2.3.1. On trouvera aussi la forme explicite des fonctions schémas $(\psi_k)_{k \in \mathbb{N}}$ ainsi que les hypothèses satisfaites par Z_k et δ_k^n, $k \in \mathbb{N}$ dans le Chapitre 2 de la Partie II.

1.6 Conclusion et perspectives

Dans cette partie de la thèse, nous nous sommes intéressés à des schémas d'approximation pour des diffusions à coefficients réguliers. L'avantage de notre approche repose sur le fait qu'elle ne traite pas le cas d'un schéma particulier mais fournit des conditions suffisantes sur le schéma d'approximation afin d'obtenir la convergence en variation totale. De plus, on suppose que les variables aléatoires utilisées pour la construction du schéma appartiennent, elles aussi, à une classe de variables aléatoires et ne suivent donc pas de loi particulière. Ainsi, les conditions (1.17) et (1.18) nous amènent à considérer le Théorème 1.5.1 comme un principe d'invariance. Par des applications directes de ce résultat, il est possible de montrer la convergence en variation totale des schéma d'Euler ou de Ninomiya Victoir par exemple et cela avec le même ordre que celui de la convergence faible pour des fonctions test régulières. De plus, une analogie avec Théorème de Prokhorov pour la convergence en variation totale du TCL, nous amène à penser que (1.17) pourrait aussi être une condition nécessaire pour obtenir la convergence dans le Théorème 1.5.1, ou du moins une piste pertinente pour en trouver une.

Une extension possible de ce résultat consiste à considérer le cas des EDS à coefficients simplement localement réguliers. Afin d'illustrer les points délicats qui apparaissent dans ce cas, nous allons nous concentrer sur le cas du processus CIR. Il s'agit d'un modèle continu unidimensionnel à valeur dans $\mathbb{R}_+$ et caractérisé par l'EDS suivante

$$dX_t \quad = (a - kX_t)dt + \sigma\sqrt{X_t}dW_t, \tag{1.34}$$

avec $a, \sigma > 0$ $k \in \mathbb{R}$. Inspiré du modèle de Vasicek (1977), qui introduit un phénomène de retour à la moyenne (lorsque $k > 0$), ce modèle a été présenté en 1985 et en modifie le coefficient de diffusion en introduisant la racine carrée, entre autre, pour garantir sa positivité. Cette propriété lui confère un intérêt particulier pour modéliser la variance d'un processus. On peut ainsi l'employer dans un modèle de covariance instantanée, comme l'a fait Heston [35]. Ce modèle est ainsi très répandu dans le monde de la finance pour modéliser des volatilités mais aussi des taux d'intérêt.
On rappelle également que pour $2a \geqslant \sigma^2$, la diffusion du CIR reste strictement positive. En revanche pour $2a < \sigma^2$, la diffusion du CIR touche 0 avec une probabilité positive et cette probabilité vaut même 1 lorsque $k \geqslant 0$. D'un point de vue heuristique, cela signifie que pour des grandes valeurs de σ^2, le CIR est susceptible de passer un temps non négligeable au voisinage de 0. Le caractère non Lipschitzien de la racine carrée dans cette zone est alors un obstacle pour la construction de schémas d'approximation ainsi que pour les preuves de convergence. Cependant, Alfonsi [3] est tout de même parvenu à introduire un schéma de discrétisation du CIR d'ordre faible $h = 2$ pour des fonctions test régulières (et à croissance polynomiale).
En reprenant ce schéma on aimerait montrer qu'il est possible d'obtenir un résultat similaire au Théorème 1.5.1 pour le CIR. Cependant nous ne sommes plus dans le cadre du Théorème 1.5.1. Une fois encore, le manque de régularité du terme de diffusion pose problème. En effet, les dérivées de la racine carrée divergent au voisinage de 0 et cela va entrainer une explosion des constantes qui apparaissent de notre estimation. Cependant, en considérant des fonctions test dont le support ne contient pas cette zone problématique, il est tout de même possible d'exploiter le Théorème 1.5.1 et ainsi espérer obtenir un résultat de convergence en variation totale pour le CIR.

Nous proposons ici une piste envisageable pour prouver cette convergence et qui pourrait s'appliquer dans le cas général des diffusions à coefficients localement bornés. En revanche, cette méthode va introduire un terme de pénalisation dans la vitesse de convergence de l'erreur. On expose les points clés de cette approche.
Tout d'abord, on considère les fonctions test f à support contenu dans $[d_1, d_2]$, $0 < d_1 < d_2 < \infty$. Lorsqu'on souhaite calculer $\mathbb{E}[f(X_T)], T > 0$, les seules réalisations du processus CIR qui nous intéressent sont donc celles telles que $X_T \in [d_1, d_2]$. Étant donné que le CIR est un processus continu, il est alors possible de trouver un compact de $\mathbb{R}_+^*$ qui contient $[d_1, d_2]$ et sur lequel ces réalisations prennent leur valeur au voisinage de T avec une probabilité très proche de 1. Comme les coefficients du CIR sont réguliers sur tout compact de $\mathbb{R}_+^*$, on pourra exploiter le Théorème 1.5.1 au voisinage de T. De plus, la probabilité que la diffusion sorte d'un tel compact au voisinage de T pourra être contrôlée par des inégalités de concentration (inégalités de Hoeffding ou Bernstein par exemple).

Afin d'identifier la vitesse de convergence qui apparaît alors, nous détaillons succinctement cette approche. On choisit δ tel que $T - \delta \in \pi_{T,n}$ et $2T/n < \delta < T$ et on prend $0 < v < d_1$. Sachant

que $X_T \in [d_1, d_2]$, il y a alors deux possibilités. Ou bien le processus reste dans $[d_1 - v, d_2 + v]$ entre $T - \delta$ et T, ou bien il en sort.
Si le processus reste dans $[d_1 - v, d_2 + v]$, on introduit des modifications régulières du CIR et de son schéma (localisées sur $[d_1 - v, d_2 + v]$) sur $[T - \delta, T]$. On peut alors utiliser une variante du Théorème 1.5.1 et il apparaît un terme d'erreur de la forme $C/(\delta^q n^h)$ où h et q sont les valeurs qui apparaissent dans $E_n(h, q)$ et $E_n^*(h, q)$ pour ces versions régulières.
A contrario, on peut contrôler la probabilité que le processus ou le schéma sorte du compact $[d_1 - v, d_2 + v]$ par le biais des inégalités de concentrations. Cette probabilité est dominée par un terme de la forme $\exp(-Cv^2/\delta)$.
Ainsi, en rassemblant ces deux estimations, pour f à support dans $[d_1, d_2]$, l'erreur faible entre le CIR (où $h = 2$ et $q = 6$) et son schéma est dominée par un terme de la forme

$$C/(\delta^q n^h)\|f\|_\infty + \exp(-Cv^2/\delta)\|f\|_\infty. \tag{1.35}$$

L'étude de cette expression va nous permettre d'établir qu'il apparaît nécessairement un terme de pénalisation dans la vitesse de convergence en variation totale. Cependant, ce terme est logarithmique et n'a donc pas une grande influence pour des applications concrètes. Pour simplifier l'approche, nous ne discuterons pas davantage du choix de d_1, d_2 et v, et on impose $v = d_1/2$.
Il convient ensuite de déterminer le choix de δ qui minimise (1.35). Si h représente l'ordre du schéma, on veut alors choisir δ tel que $\delta \leqslant Cd_1^2/(4h\ln(n))$. On a donc nécessairement $1/(\delta^q n^h) \geqslant (4h\ln(n))^q/(C^q d_1^{2q} n^h) \geqslant C\ln(n)^q/n^h$. Ainsi quelque soit le choix de δ, les contraintes imposées par les inégalités de concentration d'une part, et par la régularisation des semigroupes d'autre part impliquent la présence d'un terme de pénalisation supérieur à $\ln(n)^l$ où $l \in \mathbb{N}^*$.

Ainsi, dans le cas du CIR, on pourra espérer obtenir le résultat suivant. Soit $0 < d_1 < d_2$. Alors il existe $n_0, l \in \mathbb{N}^*, C > 0$, tels que pour toute fonction mesurable bornée f avec $\mathrm{supp}(f) \subset [d_1, d_2]$ et $n \geqslant n_0$,

$$|\mathbb{E}[f(X_T(x)) - f(X_T^n(x))]| \leqslant C(1 + |x|^\beta)\|f\|_\infty \ln(n)^l/n^2 \tag{1.36}$$

où $X(x)$ et $X^n(x)$ sont les processus partant de x, c'est à dire $X_0(x) = X_0^n(x) = x$. On pourra ensuite appliquer ce résultat dans le cas du modèle de Heston [35] et obtenir une estimation similaire.

Enfin, pour traiter exhaustivement le cas des fonctions test mesurables bornées sur $\mathbb{R}_+$, il faudrait s'intéresser à la probabilité que les valeurs finales de la diffusion et de son schéma n'appartiennent pas à $[d_1, d_2]$. Plusieurs pistes sont alors envisageables pour estimer ces probabilités. On pourra notamment étudier des estimations basées sur les densités lorsqu'elle sont connues ou encore sur les moments de tout ordre voire les moments exponentiels. Dans le cas du CIR on peut alors se demander si de telles méthodes ne permettraient pas d'obtenir (1.36) lorsque $\mathrm{supp}(f) \subset \mathbb{R}_+^*$. Cette réflexion concernant les diffusions à coefficients irréguliers n'est pas approfondie dans cette thèse mais fera l'objet de recherches ultérieures.

Chapitre 2

Estimation des paramètres du processus de Wishart

Dans la section précédente, nous avons proposé des méthodes de simulation de schémas d'approximation pour des EDS. Ces méthodes viennent en complément des méthodes de simulation exacte des processus. Ainsi, dès lors qu'on sélectionne un modèle et qu'on parvient à le simuler, l'étape suivante consiste à déterminer les valeurs des paramètres à appliquer en fonction des données réelles de la variable que l'on modélise. Dans cette section, nous nous intéressons à l'estimation des paramètres du processus de Wishart par maximum de vraisemblance. Il s'agit d'un processus à valeurs dans les matrices symétriques positives et dont la version unidimensionnelle est le CIR présenté en (1.34). Une méthode de simulation exacte de ce processus est proposée dans [1]. On contribue ici à généraliser l'estimation par maximum de vraisemblance des paramètres du CIR, qui apparaît dans [13, 14], au cas multi-dimensionnel du Wishart.

2.1 Le modèle de Wishart

Dans un premier temps, nous allons présenter le processus de Wishart. Soit $d \in \mathbb{N}^*$ la dimension, $\mathcal{M}_d$ l'ensemble des matrices carrées réelles d-dimensionnelles, $\mathcal{S}_d^+$ (resp. $\mathcal{S}_d^{+,*}$) le sous ensemble des matrices (semidéfinies) positives (resp. définies positives), $\mathcal{S}_d$ (resp. $\mathcal{A}_d$) le sous ensemble de matrices symétriques (resp. antisymétriques). Les processus de Wishart sont définis par l'EDS suivante

$$\begin{cases} dX_t = \left[\alpha a^\top a + bX_t + X_t b^\top\right] dt + \sqrt{X}_t dW_t a + a^\top dW_t^\top \sqrt{X}_t, \quad t > 0 \\ X_0 = x \in \mathcal{S}_d^+, \end{cases} \tag{2.1}$$

où $\alpha \geqslant d-1$, $a \in \mathcal{M}_d$, $b \in \mathcal{M}_d$ et $(W_t)_{t \geqslant 0}$ est une matrice de $\mathcal{M}_d$ composée de mouvements Browniens standards indépendants. On rappelle que pour $x \in \mathcal{S}_d^+$, $\sqrt{x}$ est l'unique matrice de $\mathcal{S}_d^+$ telle que $\sqrt{x}^2 = x$.
En se plaçant dans le cas plus général des diffusions affines, Bru [20] et Cuchiero et al. [21] ont montré que l'EDS Equation (2.1) admet une unique solution forte lorsque $\alpha \geqslant d+1$ et une unique solution faible pour $\alpha \geqslant d-1$. De plus, $X_t \in \mathcal{S}_d^{+,*}$ pour tout $t \geqslant 0$ lorsque $x \in \mathcal{S}_d^{+,*}$ et $\alpha \geqslant d+1$. On appliquera alors la notation $WIS_d(x, \alpha, b, a)$ pour désigner la loi de $(X_t)_{t \geqslant 0}$ et $WIS_d(x, \alpha, b, a; t)$ pour celle de X_t.
Ce modèle a été introduit par Bru [19] afin de traiter des données biologiques. A l'instar du CIR, le processus de Wishart présente de nombreuses propriétés qui le rendent à la fois pertinent pour le type de variable qu'il peut modéliser mais aussi pratique pour l'étude analytique puisqu'il s'agit

d'un modèle affine. Premièrement, on rappelle que pour α entier, le processus de Wishart n'est autre qu'un processus d'Ornstein Uhlenbeck matriciel multiplié par sa transposée. Dans le cas unidimensionnel, le CIR peut lui aussi être associé au carré d'un processus de Bessel. Un de ces intérêts principaux repose sur le fait qu'il prenne ses valeur dans $\mathcal{S}_d^+$. En effet, cette propriété lui confère la possibilité d'être utilisé pour modéliser une matrice de covariance et non plus seulement une variance. Ainsi, il permet entre autres, la généralisation du modèle d'Heston [35] au cas multi-dimensionnel d'un panier d'actifs. Ce dernier s'écrit

$$\begin{cases} d\log(S_t) = (\beta + (\mathrm{Tr}[\gamma_1\sqrt{X_t}], \ldots, \mathrm{Tr}[\gamma_d\sqrt{X_t}]))dt + \sigma\sqrt{X_t}dB_t, & S_0 \in (0,\infty)^d, \\ dX_t = \left[\alpha a^\top a + bX_t + X_t b^\top\right] dt + \sqrt{X}_t dW_t a + a^\top dW_t^\top \sqrt{X}_t, & X_0 = x \in \mathcal{S}_d^+, \end{cases} \tag{2.2}$$

où $\beta \in \mathbb{R}^d$, $\gamma_1, ., \gamma_d \in \mathcal{M}_d$, $\sigma \in \mathcal{S}_d^{+,*}$ et B_t est un mouvement Brownien d-dimensionnel indépendant de W_t. De plus, les paramètres de la composante Wishart vérifient les hypothèses mentionnées plus haut. De cette façon, on peut modéliser une variable d-dimensionnelle strictement positive (ici S), dont la covariance instantanée entre les composantes est elle aussi un processus aléatoire (ici X). Cette approche a été utilisée par Gourieroux et Sufana [32] ainsi que par Da Fonseca et al. [23] qui l'ont employée pour modéliser la matrice de covariance instantanée entre actifs financiers, dans un modèle de covariance stochastique similaire à (2.2). Ils traitent ainsi le cas d'un panier d'actifs. Tout comme le CIR, les processus de Wishart sont aussi utilisés dans les modèles de taux d'intérêt. En effet, lorsque $-b \in \mathcal{S}_d^{+,*}$, on observe un phénomène de retour à la moyenne similaire à celui du CIR et particulièrement adapté aux besoins de ce domaine.
De plus, il s'agit de processus affines et il est donc possible de recourir à des calculs analytiques pour obtenir certaines propriétés. Ainsi, en se ramenant à des équations de Riccati, il est par exemple possible de calculer ses transformées de Fourier ou de Laplace et d'en déduire des calculs d'espérances explicites. Par le biais de méthodes numériques, il est alors possible de mettre en œuvre des procédés d'inversion de ces transformées et d'obtenir des calculs de prix de produits financiers. Des processus affines impliquant le modèle de Wishart ont été proposés par Gourieroux et Sufana [33], Gnoatto [28] ainsi qu'Ahdida et al. [2].

Une telle ampleur dans les utilisations possibles de ce modèle nécessite donc une estimation précise de ses paramètres. Ce sujet a déjà été abordé par Da Fonseca et al. [22] pour le modèle présenté dans [23]. Cependant, aucune méthode proposée jusqu'à maintenant ne traite le cas de l'estimation par maximum de vraisemblance. On précisera que dans le cas du CIR, l'estimation de paramètres par maximum de vraisemblance a été étudiée par Ben Alaya and Kebaier [13,14]. Dans ce travail, on généralise donc ces résultats dans le cas multi-dimensionnel.

2.2 Présentation du problème

On rappelle tout d'abord que la loi de X ne dépend de a qu'à travers $a^\top a$ étant donné que

$$WIS_d(x, \alpha, b, a) \underset{loi}{=} WIS_d(x, \alpha, b, \sqrt{a^\top a}),$$

voir par exemple l'équation (12) dans [1]. Afin de mettre en pratique le modèle de Wishart, il y a donc trois paramètres à estimer : α, b et $a^\top a$. Dans cette étude, on s'appuie sur la théorie développée dans Lipster et Shiryaev [50] et Kutoyants [45] et on suppose qu'on observe la trajectoire $(X_t, t \in [0,T])$ jusqu'au temps $T > 0$. D'un point de vue mathématique, cette approche s'avère très pratique. En revanche, elle reste irréaliste dans le cas des applications concrètes où les observations sont discrètes. Cet aspect n'est pas abordé au cours de cette thèse.

Cependant, par une étude numérique, on observe que l'approximation du processus sur une grille de temps fournit une approximation satisfaisante des estimateurs.
Tout d'abord, on remarque que si la trajectoire $(X_t, t \in [0,T])$ est connue alors on peut connaître exactement la valeur de $a^\top a$ en exploitant la covariation quadratique du processus (voir par exemple le Lemme 2 dans [1]). En effet, pour $i,j,k,l \in \{1,\dots,d\}$, on a

$$\langle X_{i,j}, X_{k,l}\rangle_T = \int_0^T (a^\top a)_{j,l}(X_s)_{i,k} + (a^\top a)_{j,k}(X_s)_{i,l} + (a^\top a)_{i,l}(X_s)_{j,k} + (a^\top a)_{i,k}(X_s)_{j,l} ds. \quad (2.3)$$

Par conséquent,

$$\begin{aligned} (a^\top a)_{i,i} &= \frac{1}{4}\langle X_{i,i}\rangle_T \Big(\int_0^T (X_s)_{i,i} ds\Big)^{-1}, \quad (2.4)\\ (a^\top a)_{i,j} &= \left(\frac{1}{2}\langle X_{i,j}, X_{i,i}\rangle_T - (a^\top a)_{i,i}\int_0^T (X_s)_{i,j} ds\right)\Big(\int_0^T (X_s)_{i,i} ds\Big)^{-1}, \end{aligned}$$

pour $1 \leqslant i,j \leqslant d$ et $j \neq i$. En utilisant par exemple la Proposition 4 dans [20], on déduit que la covariation quadratique de $(X_t, t \in [0,T])$ est finie et que $X_t \in \mathcal{S}_d^{+,*}$ dt-p.p. Ainsi les quantités introduites dans (2.4) sont bien définies. On supposera alors que $a^\top a \in \mathcal{S}_d^{+,*}$ et on notera par $a \in \mathcal{M}_d$ une matrice inversible dont le carré correspond aux observations de $a^\top a$. Une méthode possible consiste à prendre la matrice de Cholevsky de $a^\top a$. C'est le choix qui a été fait pour les approximation numériques. De plus, on rappelle que $Y_t = (a^\top)^{-1} X_t a^{-1}$ suit la loi $WIS_d((a^\top)^{-1} x a^{-1}, \alpha, (a^\top)^{-1} b a^\top, I_d)$, voir par exemple l'équation (13) dans [1]. Cette transformation nous permet de simplifier notre problème. En effet, la connaissance de la trajectoire nous permet d'estimer exactement a et ainsi notre problème revient à estimer α et b dans le cas $a = I_d$.

2.3 Estimateurs du maximum de vraisemblance des paramètres du processus de Wishart

Nous allons maintenant présenter l'EMV de $\theta = (\alpha, b)$ dans le cas $a = I_d$. On notera par $\mathbb{P}_\theta$ la probabilité de référence sous laquelle X satisfait l'EDS

$$dX_t = \Big[\alpha I_d + bX_t + X_t b^\top\Big] dt + \sqrt{X_t} dW_t + dW_t^\top \sqrt{X_t}. \quad (2.5)$$

On introduit également $\alpha_0 \geqslant d+1$ et $\theta_0 = (\alpha_0, 0)$. Pour l'estimation jointe de α et b on supposera que

$$\alpha \geqslant d+1 \text{ et } x \in \mathcal{S}_d^{+,*}. \quad (2.6)$$

En utilisant le Theorem 4.1 dans Mayerhofer [54], on sait alors que

$$\frac{d\mathbb{P}_{\theta_0,T}}{d\mathbb{P}_{\theta,T}} := \exp\left(\int_0^T \mathrm{Tr}[H_s dW_s] - \frac{1}{2}\int_0^T \mathrm{Tr}[H_s H_s^\top] ds\right), \text{ avec } H_t = \frac{\alpha_0 - \alpha}{2}(\sqrt{X_t})^{-1} - b\sqrt{X_t},$$

où $\mathbb{P}_{\theta,T}$ est la restriction $\mathbb{P}_\theta$ à la tribu $\sigma(W_s, s \in [0,T])$, définie un changement de probabilité. De plus, sous $\mathbb{P}_{\theta_0,T}$, $\tilde{W}_t = W_t - \int_0^t H_s^\top ds$ est un mouvement brownien standard $d \times d$-dimensionnel. On a alors

$$dX_t = \alpha_0 I_d dt + \sqrt{X_t} d\tilde{W}_t + d\tilde{W}_t^\top \sqrt{X_t},$$

et ainsi X est un processus de Wishart ayant pour paramètres θ_0 sous $\mathbb{P}_{\theta_0}$. Réciproquement, le même Théorème implique que

$$\begin{aligned}
\frac{d\mathbb{P}_{\theta,T}}{d\mathbb{P}_{\theta_0,T}} &= \exp\left(-\int_0^T \operatorname{Tr}[H_s d\tilde{W}_s] - \frac{1}{2}\int_0^T \operatorname{Tr}[H_s H_s^\top]ds\right) \\
&= \exp\left(\frac{\alpha-\alpha_0}{2}\int_0^T \operatorname{Tr}[(\sqrt{X_s})^{-1}d\tilde{W}_s] + \int_0^T \operatorname{Tr}[b\sqrt{X_s}d\tilde{W}_s]\right. \\
&\quad \left. - \frac{(\alpha-\alpha_0)^2}{8}\int_0^T \operatorname{Tr}[X_s^{-1}]ds - \frac{1}{2}\int_0^T \operatorname{Tr}[bX_s b^\top]ds - \frac{(\alpha-\alpha_0)T}{2}\operatorname{Tr}[b]\right).
\end{aligned} \tag{2.7}$$

est aussi un changement de probabilité. On en déduit que les mesures de probabilité $\mathbb{P}_{\theta,T}$ et $\mathbb{P}_{\theta_0,T}$ sont équivalentes. Afin d'être interprété comme une vraisemblance par rapport au processus de Wishart, le membre de droite dans Equation (2.7) doit pouvoir être exprimé en fonction de la trajectoire $(X_t, t \in [0,T])$. Malheureusement, cette propriété n'est garantie que si b est une matrice symétrique. C'est une conséquence directe du résultat suivant.

Proposition 2.3.1. *Soit $(\mathcal{F}_t^X)_{t\geqslant 0}$ la filtration engendrée par le processus X. Alors, $\frac{d\mathbb{P}_{\theta,T}}{d\mathbb{P}_{\theta_0,T}} \in \mathcal{F}_T^X \iff b \in \mathcal{S}_d$, et dans ce cas $\frac{d\mathbb{P}_{\theta,T}}{d\mathbb{P}_{\theta_0,T}} = L_T^{\theta,\theta_0}$ où*

$$\begin{aligned}
L_T^{\theta,\theta_0} = \exp\Big(&\frac{\alpha-\alpha_0}{4}\log\left(\frac{\det[X_T]}{\det[x]}\right) + \frac{\operatorname{Tr}[bX_T]-\operatorname{Tr}[bx]}{2} - \frac{1}{2}\int_0^T \operatorname{Tr}[b^2X_s]ds \\
&- \frac{\alpha-\alpha_0}{4}\left(\frac{\alpha+\alpha_0}{2}-1-d\right)\int_0^T \operatorname{Tr}[X_s^{-1}]ds - \frac{\alpha T}{2}\operatorname{Tr}[b]\Big).
\end{aligned} \tag{2.8}$$

Désormais, on suppose donc que

$$b \in \mathcal{S}_d. \tag{2.9}$$

Notons que cette hypothèse ne s'avère pas extrêmement restrictive pour des applications concrètes. En effet, dans la majorité des cas, on souhaite modéliser des processus ergodiques et ainsi on a $-b \in \mathcal{S}_d^{+,*}$.

Maintenant, on observe que la quantité qui apparaît à l'intérieur de l'exponentielle dans Equation (2.8) est quadratique par rapport à (α, b) et diverge presque sûrement vers $-\infty$ lorsque $\|(\alpha,b)\| \to +\infty$. On en déduit qu'il existe un unique maximum global à Equation (2.8) sur $\mathbb{R}\times\mathcal{S}_d$. L'EMV $\hat\theta_T = (\hat\alpha_T, \hat b_T)$ de (α, b) est alors donné par ce maximum et vérifie la formule suivante

$$\begin{cases}
\frac{1}{4}\log\left(\frac{\det[X_T]}{\det[x]}\right) - \frac{\hat\alpha_T-1-d}{4}\int_0^T \operatorname{Tr}[X_s^{-1}]ds - \frac{T}{2}\operatorname{Tr}[\hat b_T] = 0, \\
\frac{X_T-x}{2} - \frac{1}{2}\int_0^T(\hat b_T X_s + X_s\hat b_T)ds - \frac{\hat\alpha_T T}{2}I_d = 0.
\end{cases} \tag{2.10}$$

En inversant ce système linéaire, nous allons ensuite donner une forme explicite pour $(\hat\alpha_T, \hat b_T)$. Pour cela on commence par introduire les applications linéaires

$$\begin{aligned}
\mathcal{L}_X :\ & \mathcal{S}_d \to \mathcal{S}_d & \text{et } \mathcal{L}_{X,a} :\ & \mathcal{S}_d \to \mathcal{S}_d \\
& Y \mapsto YX + XY & & Y \mapsto YX + XY - 2a\operatorname{Tr}[Y]I_d.
\end{aligned} \tag{2.11}$$

où $X \in \mathcal{S}_d$ and $a \in \mathbb{R}$. Afin de simplifier les écritures, nous adopterons les notations suivantes :

$$R_T := \int_0^T X_s ds, \quad Q_T := \left(\int_0^T \operatorname{Tr}[X_s^{-1}]ds\right)^{-1}, \quad Z_T := \log\left(\frac{\det[X_T]}{\det[x]}\right), \tag{2.12}$$

et on remarque que les variables Q_T et Z_T sont bien définies si $\alpha \geqslant d+1$ tandis que R_T est bien défini pour $\alpha \geqslant d-1$ et appartient à $\mathcal{S}_d^{+,*}$ presque sûrement. En réécrivant Equation (2.10), on a $\hat{\alpha}_T = 1+d+Q_T(Z_T - 2T\operatorname{Tr}[\hat{b}_T])$ et $\mathcal{L}_{R_T,T^2Q_T}(\hat{b}_T) = X_T - x - T\,(Q_TZ_T + 1 + d)\,I_d$. On peut montrer que les applications linéaires en jeu peuvent être inversées dans notre cas et on conclut que

$$\begin{cases} \hat{\alpha}_T &= 1+d+Q_T\Big(Z_T - 2T\operatorname{Tr}\left[\mathcal{L}^{-1}_{R_T,T^2Q_T}\left(X_T - x - T\left[Q_TZ_T+1+d\right]I_d\right)\right]\Big) \\ \hat{b}_T &= \mathcal{L}^{-1}_{R_T,T^2Q_T}\left(X_T - x - T\left[Q_TZ_T + 1 + d\right]I_d\right). \end{cases} \tag{2.13}$$

On note que pour $\alpha \in [d-1, d+1)$, alors l'EMV de α n'est plus défini. Le même phénomène apparaît déjà dans le cas unidimensionnel du CIR comme souligné par Ben Alaya and Kebaier [13]. Cependant il est toujours possible d'estimer b si $\alpha \geqslant d-1$ est connu. On utilise alors les notations $\theta = (\alpha, b)$ et $\theta_0 = (\alpha, 0)$. Par une méthode similaire à celle de l'estimation du couple on a

$$\frac{d\mathbb{P}_{\theta,T}}{d\mathbb{P}_{\theta_0,T}} = \exp\left(\int_0^T \operatorname{Tr}[b\sqrt{X_s}d\tilde{W}_s] - \frac{1}{2}\int_0^T \operatorname{Tr}[b^2X_s]ds\right).$$

La vraisemblance et le MLE sont alors donnés par

$$L_T^{\theta,\theta_0} = \exp\Big(\frac{\operatorname{Tr}[bX_T] - \operatorname{Tr}[bx]}{2} - \frac{1}{2}\int_0^T \operatorname{Tr}[b^2X_s]ds - \frac{\alpha T}{2}\operatorname{Tr}[b]\Big), \tag{2.14}$$

$$\hat{b}_T = \mathcal{L}^{-1}_{R_T}\left(X_T - x - \alpha T I_d\right). \tag{2.15}$$

Par un résultat tout à fait semblable à la Proposition 2.3.1, l'hypothèse $b \in \mathcal{S}_d$ est à nouveau nécessaire (et suffisante) pour obtenir $\frac{d\mathbb{P}_{\theta,T}}{d\mathbb{P}_{\theta_0,T}} \in \mathcal{F}_T^X$. On peut maintenant établir les résultats de convergence.

2.4 Convergence des estimateurs

Dans cette section, on va énoncer les vitesses de convergence des EMV dans des cas ergodiques et non ergodiques. On identifie aussi les lois limites vers lesquelles convergent les EMV renormalisés par leur vitesse. On remarquera que dans le cas $d=1$ on obtient les mêmes résultats que dans [13] pour le CIR.

Cas ergodique
On rappelle que $-b \in \mathcal{S}_d^{+,*}$ dans ce cas. Avant d'énoncer les résultats de convergence, on introduit les constantes suivantes

$$\overline{R}_\infty := \mathbb{E}_\theta(X_\infty) = -\frac{\alpha}{2}b^{-1} \in \mathcal{S}_d^{+,*}, \quad \overline{Q}_\infty := \frac{1}{\mathbb{E}_\theta(\operatorname{Tr}[X_\infty^{-1}])} = \frac{\alpha - (1+d)}{2\operatorname{Tr}[-b]}. \tag{2.16}$$

Dans le cas ou $\alpha \geqslant d+1$, l'estimateur du couple est bien défini et il converge.

Théorème 2.4.1. *Soit $X \sim WIS_d(x, \alpha, b, I_d)$ avec $-b, x \in S_d^{+,*}$, $\alpha \geqslant d+1$. Pour $T > 0$, on considère l'estimateur $(\hat{\alpha}_T, \hat{b}_T)$ de (α, b) défini par (2.13). On a les propriétés suivantes*

A. *On suppose que $\alpha > d+1$. Alors, $\left(\sqrt{T}(\hat{b}_T - b, \hat{\alpha}_T - \alpha)\right)$ converge en loi sous $\mathbb{P}_\theta$, lorsque $T \to +\infty$, vers le vecteur Gaussien centré $(\mathbf{G}, H)$ à valeurs dans $\mathcal{S}_d \times \mathbb{R}$ et dont la transformée de Laplace est donnée par*

$$c, \lambda \in \mathcal{S}_d \times \mathbb{R}, \quad \mathbb{E}_\theta\left[\exp\left(\operatorname{Tr}[c\mathbf{G}] + \lambda H\right)\right] = \\ \exp\left(\frac{2\overline{Q}_\infty \lambda^2}{1 - \overline{Q}_\infty \operatorname{Tr}[\overline{R}_\infty^{-1}]} - \frac{2\overline{Q}_\infty \lambda}{1 - \overline{Q}_\infty \operatorname{Tr}[\overline{R}_\infty^{-1}]}\operatorname{Tr}[c\overline{R}_\infty^{-1}] + \operatorname{Tr}[c\mathcal{L}^{-1}_{\overline{R}_\infty,\overline{Q}_\infty}(c)]\right).$$

B. *On suppose que* $\alpha = d+1$. *Alors,* $\left(\sqrt{T}(\hat{b}_T - b), T(\hat{\alpha}_T - \alpha)\right)$ *converge en loi sous* $\mathbb{P}_\theta$*, lorsque* $T \to +\infty$*, vers* $\left(\mathbf{G}, -2\tau^{-1}_{-\operatorname{Tr}[b]} \operatorname{Tr}[b+\mathbf{G}]\right)$*, où* $\tau_a = \inf\{t \geq 0,\ B_t = a\}$ *avec* $(B_t)_{t\geq 0}$ *un mouvement Brownien standard unidimensionnel et* $\mathbf{G}$ *est un vecteur Gaussien indépendant de* B *tel que* $\mathbb{E}_\theta\left[\exp\left(\operatorname{Tr}[c\mathbf{G}]\right)\right] = \exp\left(\operatorname{Tr}[c\mathcal{L}^{-1}_{\overline{R}_\infty}(c)]\right)$*,* $c \in \mathcal{S}_d$.

Lorsque $d-1 \leqslant \alpha < d+1$, l'estimateur du couple n'est plus défini. Il en serait de même si on traitait le cas de l'estimateur de α seulement. En revanche, l'estimateur de b seul reste bien défini et on obtient la même vitesse de convergence que pour le couple.

Théorème 2.4.2. *Soit* $X \sim W_d(x, \alpha, b, I_d)$ *avec* $-b, x \in S_d^{+,*}$ *et* $\alpha \geqslant d-1$*. Pour* $T > 0$*, on considère l'estimateur* $\hat{b}_T$ *de* b *défini par (2.15). Alors* $\sqrt{T}(\hat{b}_T - b)$ *converge en loi sous* $\mathbb{P}_\theta$ *vers le vecteur Gaussien centré* $\mathbf{G}$*, dont la transformée de Laplace est donnée par :* $\mathbb{E}_\theta\left[\exp\left(\operatorname{Tr}[c\mathbf{G}]\right)\right] = \exp\left(\operatorname{Tr}[c\mathcal{L}^{-1}_{\overline{R}_\infty}(c)]\right)$*,* $c \in \mathcal{S}_d$.

On traite donc complètement le cas ergodique pour les EMV de (α, b) et de b. Il s'agit du cas le plus largement utilisé pour des applications concrètes. En effet, rares sont les cas où on souhaite mettre en œuvre des modèles qui divergent en temps long. Cependant, on aborde tout de même certains cas non ergodiques afin de compléter cette étude mathématique de la manière la plus exhaustive possible.

Cas non ergodique

Nous nous intéressons maintenant au cas $-b \notin \mathcal{S}_d^{+,*}$. Ici, nous ne serons pas en mesure de traiter complètement cette configuration mais nous démontrerons certains cas de convergence des EMV. On observera que les vitesses d'estimation sont différentes de celles du cas ergodique, plus rapide pour b et plus lente pour α. Dans le cas de l'EMV du couple (α, b), nous obtenons la convergence si $b = 0$.

Théorème 2.4.3. *Soit* $X \sim W_d(x, \alpha, 0, I_d)$ *avec* $x \in S_d^{+,*}$*,* $\alpha \geqslant d+1$*. Pour* $T > 0$*, on considère l'estimateur* $(\hat{\alpha}_T, \hat{b}_T)$ *de* $(\alpha, 0)$ *défini par (2.13). On a les propriétés suivantes*

A. *On suppose que* $\alpha > d+1$*. Alors,* $(T(\hat{b}_T - b), \sqrt{\log(T)}(\hat{\alpha}_T - \alpha))$ *converge en loi sous* $\mathbb{P}_\theta$*, lorsque* $T \to +\infty$*, vers*

$$\left(\mathcal{L}^{-1}_{R_1^0}\left(X_1^0 - \alpha I_d\right), 2\sqrt{\frac{\alpha-(d+1)}{d}}G\right),$$

où $X_t^0 = \alpha t I_d + \int_0^t \sqrt{X_s^0}dW_s + dW_s^\top \sqrt{X_s^0} \sim WIS_d(0, \alpha, 0, I_d)$*,* $R_t^0 = \int_0^t X_s^0 ds$ *et* $G \sim \mathcal{N}(0,1)$ *est une variable aléatoire normale centrée réduite indépendante.*

B. *On suppose que* $\alpha = d+1$*. Alors,* $(T(\hat{b}_T - b), \log(T)(\hat{\alpha}_T - \alpha))$ *converge en loi sous* $\mathbb{P}_\theta$*, lorsque* $T \to +\infty$*, vers*

$$\left(\mathcal{L}^{-1}_{R_1^0}\left(X_1^0 - \alpha I_d\right), \frac{4}{d\tau_1}\right),$$

où $X_t^0 = \alpha t I_d + \int_0^t \sqrt{X_s^0}dW_s + dW_s^\top \sqrt{X_s^0} \sim WIS_d(0, \alpha, 0, I_d)$*,* $R_t^0 = \int_0^t X_s^0 ds$ *et* $\tau_1 = \inf\{t \geqslant 0, B_t = 1\}$ *avec* B *un mouvement Brownien standard indépendant de* W.

Il est important de remarquer que dans ce cas, la vitesse de convergence de la composante α s'avère très voire trop lente pour des applications concrètes ce qui n'est pas le cas de la composante b.

Nous nous intéressons maintenant à la convergence de l'estimateur de b lorsque $\alpha \geqslant d-1$ est connu. Ici nous allons être capable de traiter davantage de configurations que pour l'estimation du couple. Plus particulièrement, nous allons identifier les vitesses de convergence et les lois limites de l'estimateur de b lorsque $b = b_0 I_d$ pour $b_0 = 0$ et pour $b_0 > 0$. Dans ce dernier cas, qui n'est pas abordé pour l'estimateur du couple, la vitesse de convergence est exponentielle ce qui se révèle relativement pratique pour des applications sur des données réelles.

Théorème 2.4.4. *Soit $X \sim W_d(x, \alpha, b, I_d)$ avec $x \in S_d^{+,*}$ et $\alpha \geqslant d-1$. Pour $T > 0$, on considère l'estimateur $\hat{b}_T$ de b défini par (2.15).*

A. *On suppose que $b = 0$ et $\alpha \geqslant d-1$. Alors $T(\hat{b}_T - b)$ converge en loi sous $\mathbb{P}_\theta$ vers $\mathcal{L}_{R_1^0}^{-1}\left(X_1^0 - \alpha I_d\right)$, où $X_t^0 = \alpha t I_d + \int_0^t \sqrt{X_s^0} dW_s + dW_s^\top \sqrt{X_s^0} \sim WIS_d(0, \alpha, 0, I_d)$ et $R_t^0 = \int_0^t X_s^0 ds$.*

B. *On suppose que $b = b_0 I_d$, $b_0 > 0$ et $\alpha \geqslant d-1$. Alors $\exp(b_0 T)(\hat{b}_T - b)$ converge en loi sous $\mathbb{P}_\theta$ vers $\mathcal{L}_X^{-1}\left(\sqrt{X}\tilde{\mathbf{G}} + \tilde{\mathbf{G}}\sqrt{X}\right)$ où $X \sim WIS_d\left(\frac{x}{2b_0}, \alpha, 0, I_d; \frac{1}{4b_0^2}\right)$ et $\tilde{\mathbf{G}}$ est une matrice aléatoire de $\mathcal{M}_d$ indépendante de X et dont les composantes suivent des lois normales centrées réduites indépendantes.*

En l'état, nous ne sommes pas en mesure de traiter le cas général $b \in \mathcal{S}_d$. Le point problématique repose sur le fait que la renormalisation par la vitesse de $\hat{b}_t - b$ n'est plus scalaire mais matricielle. En effet, lorsque $-b \in \mathcal{S}_d^{+,*}$, $b = 0$ ou $b = b_0 I_d$, la vitesse de convergence est la même pour tous les termes de la matrice $\hat{b}_t - b$. Dans le cas général $b \in \mathcal{S}_d$, on peut conjecturer que la vitesse de chaque composante va être un mélange complexe de ces trois types de vitesse dû aux multiples produits matriciels qui apparaissent lors de l'estimation. Afin d'illustrer cette difficulté, on traite le cas de l'estimateur de b lorsqu'on sait que b est diagonale. Pour plus de détails, on invite le lecteur à se reporter à la Section 1.3.2 de la Partie III.

2.5 La transformée de Laplace de (X_T, R_T)

Dans cette Section, on présente le résultat principal concernant le transformée de Laplace jointe de (X_T, R_T). Cette transformée a été étudiée par Bru [20], équation (4.7) lorsque $b = 0$ et plus récemment calculée explicitement par Gnoatto et Grasselli [29]. Nous contribuons à étendre leurs résultats. D'une part, on montre que son expression est valable pour tout $b \in \mathcal{S}_d$ et $\alpha \geqslant d-1$. De plus, on étend le domaine de définition de cette transformée établi dans [29]. Ces résultats ne sont pas seulement esthétiques mais s'avèrent primordiaux dans notre étude pour traiter des cas non ergodiques.

Proposition 2.5.1. *Soit $\alpha \geqslant d-1$, $x \in \mathcal{S}_d^+$, $b \in \mathcal{S}_d$ et $X \sim WIS_d(x, \alpha, b, I_d)$. Soit $v, w \in \mathcal{S}_d$ tels que*

$$\exists m \in \mathcal{S}_d,\ \frac{v}{2} - mb - bm - 2m^2 \in \mathcal{S}_d^+ \text{ et } \frac{w}{2} + m \in \mathcal{S}_d^+. \tag{2.17}$$

Alors, pour tout $t \geqslant 0$

$$\begin{aligned} &\mathbb{E}\Big[\exp\Big(-\frac{1}{2}\operatorname{Tr}[wX_t] - \frac{1}{2}\operatorname{Tr}[vR_t]\Big)\Big] \\ &= \frac{\exp\left(-\frac{\alpha}{2}\operatorname{Tr}[b]t\right)}{\det[V_{v,w}(t)]^{\frac{\alpha}{2}}}\exp\Big(-\frac{1}{2}\operatorname{Tr}\left[(V'_{v,w}(t)V_{v,w}(t)^{-1} + b)x\right]\Big), \end{aligned} \tag{2.18}$$

avec

$$V_{v,w}(t) = \left(\sum_{k=0}^\infty t^{2k+1}\frac{\tilde{v}^k}{(2k+1)!}\right)\tilde{w} + \sum_{k=0}^\infty t^{2k}\frac{\tilde{v}^k}{(2k)!}, \quad \tilde{v} = v + b^2, \quad \text{and} \quad \tilde{w} = w - b.$$

De plus, si $\tilde{v} = v + b^2 \in \mathcal{S}_d^{+,*}$*, on a* $V_{v,w}(t) = (\sqrt{\tilde{v}})^{-1}\sinh(\sqrt{\tilde{v}}t)\tilde{w} + \cosh(\sqrt{\tilde{v}}t)$ *et alors* $V'_{v,w}(t) = \cosh(\sqrt{\tilde{v}}t)\tilde{w} + \sinh(\sqrt{\tilde{v}}t)\sqrt{\tilde{v}}$.

2.6 Conclusion et perspectives

Dans cette partie, nous avons présenté une méthode d'estimation des paramètres $a^\top a$, α et b pour une diffusion $X \sim WIS_d(x, \alpha, b, a)$ dont on observe la trajectoire sur un intervalle $[0, T]$, $T > 0$. Dans ce cas, $a^\top a$ est connu et on se ramène à l'étude des EMV de α et b pour $a = I_d$. Notre démarche permet de traiter la convergence de ces estimateurs (lorsqu'ils sont bien définis) dans le cas ergodique et dans certains cas non ergodiques. Pour des applications concrètes, le cas ergodique est de loin le plus répandu et notre étude traite exhaustivement cette approche. D'ailleurs, dans les cas non ergodiques traités la convergence de la composante α s'avère trop lente pour obtenir des résultats pertinents avec un horizon de données raisonnable. A contrario, l'estimation de b est plus rapide et se prête donc bien à l'analyse de données réelles. En revanche pour être complet au sujet de l'étude mathématique du modèle, certains aspects restent encore à être aborder.

Tout d'abord, l'étude sur l'EMV n'est pas complète et ne traite pas certains cas non ergodiques ($-b \notin \mathcal{S}_d^{+,*}$). Plus particulièrement, lorsque les valeurs propres de b sont différentes, les vitesses de convergence termes à termes de $\hat{b}_t - b$ semblent l'être aussi. Cette intuition est confirmée par notre étude numérique bien qu'on ne soit pas parvenu à identifier clairement la bonne normalisation.

De plus, notre analyse mathématique de l'EMV repose sur l'hypothèse selon laquelle on observe la trajectoire du processus de Wishart en temps continu. Or, une méthode réaliste pour des applications impose qu'on ne connaisse les valeurs du processus seulement sur une grille de temps. Bien que l'étude numérique montre que cette approximation n'altère pas la convergence des estimateurs, ni leur vitesse, pour des tailles "raisonnables" de la grille, une étude complète est nécessaire pour connaître son réel impact. On peut par exemple s'inspirer d'Alaya et Kebaier [14] qui traitent le cas unidimensionnel du CIR.

Enfin, on rappelle que l'EMV n'est pas défini lorsque $b \notin \mathcal{S}_d$. Dans ce cas, il faut donc utiliser une nouvelle approche. Une piste repose sur le fait que la loi jointe du processus de Wishart à différents instants dépend de la partie antisymétrique de b. Ainsi, en utilisant l'estimateur des moments par exemple, et en étudiant des processus de la forme $t^{-1}\int_0^t f(X_s, X_{s+1})ds$, $t > 0$, pour f polynomiale, on peut espérer estimer cette partie antisymétrique.

Part II

Total variation convergence for approximation schemes

Chapter 1

A general result for total variation convergence of approximation scheme

Ce Chapitre est un article écrit avec V.Bally actuellement en soumission [10].

Abstract

In this paper, we consider Markov chains of the form $X^n_{(k+1)/n} = \psi_k(X^n_{k/n}, Z_{k+1}/\sqrt{n}, 1/n)$ where the innovation comes from the sequence $Z_k, k \in \mathbb{N}^*$ of independent centered random variables with arbitrary law. Then, we study the convergence $\mathbb{E}[f(X^n_t)] \to \mathbb{E}[f(X_t)]$ where $(X_t)_{t\geqslant 0}$ is a Markov process in continuous time. This may be considered as an invariance principle, which generalizes the classical Central Limit Theorem to Markov chains. Alternatively (and this is the main motivation of our paper), X^n may be an approximation scheme used in order to compute $\mathbb{E}[f(X_t)]$ by Monte Carlo methods. Estimates of the error are given for smooth test functions f as well as for measurable and bounded f. In order to prove convergence for measurable test functions we assume that Z_k satisfies Doublin's condition and we use Malliavin calculus type integration by parts formulas based on the smooth part of the law of Z_k. As an application, we will give estimates of the error in total variation distance for the Ninomiya Victoir scheme.

1.1 Introduction

In this paper, we consider a time grid $t^n_k = k/n, k \in \mathbb{N}$ with $n \in \mathbb{N}^*$ and a Markov chain

$$X^n_{t^n_{k+1}} = \psi_k(X^n_{t^n_k}, Z_{k+1}/\sqrt{n}, 1/n),$$

where $\psi_k : \mathbb{R}^d \times \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}^d$ is a smooth function and $Z_k, k \in \mathbb{N}^*$, is a sequence of independent centered random variables. We aim to study the convergence of the law of X^n to the law of a Markov process X. More precisely, we will give estimates of the weak error

$$\varepsilon_n(f) = |\mathbb{E}[f(X^n_t)] - \mathbb{E}[f(X_t)]| \, .$$

This problem may be considered from two points of view. The first one is to look at this convergence result as to an invariance principle. We illustrate this approach with the Central Limit Theorem (CLT). Indeed, if $\psi_k(x, z, t) = x + z$ and $Z_k, k \in \mathbb{N}^*$, are independent and identically distributed with variance 1, we have $X^n_1 = n^{-1/2}\sum_{k=1}^n Z_k$. Using then the CLT, we know that

$X_1^n \overset{law}{\to} W_1$ where $(W_t)_{t\geqslant 0}$ is a standard Brownian motion and then $W_1 \sim \mathcal{N}(0,1)$ where $\mathcal{N}(0,1)$ is the standard Normal distribution. Since the law of $Z_k, k \in \mathbb{N}^*$ is arbitrary and the limit law of $(X_1^n)_{n\in\mathbb{N}}$ does not depend on this law, we say that it is an invariance principle. Keeping this in mind, we look at our Markov chain X^n as to a general Markovian scheme based on the sequence of random variables $Z_k, k \in \mathbb{N}^*$. Then, the convergence of X^n to a Markov process X which is universal (in the sense that it does not depend on the law of $Z_k, k \in \mathbb{N}^*$) represents an invariance principle. Our result can thus be seen as a direct generalization of the CLT. Notice that, when looking from this point of view, $\psi_k, k \in \mathbb{N}$ represents a scheme which naturally appears in a concrete modelization problem. A main interest is to approximate the law of X_1^n, which may be difficult to understand directly, by the law of X_1 which is simpler to study (as for W_1 above).

A second point of view comes from numerical probabilities: For instance, if X is a diffusion process and we want to compute $\mathbb{E}[f(X_t)]$, then we can use a discretization scheme X^n (for example the Euler scheme). Thereafter, we can obtain the approximation $\mathbb{E}[f(X_t^n)]$ using Monte Carlo methods. In this kind of approaches, we may choose the approximation scheme $(X^n_{t^n_k})_{k\in\mathbb{N}}$ as we want (in contrast with the previous situation when the Markov chain X^n was given by an external modelization).

Our initial motivation for the study of the error $\varepsilon_n(f)$ comes from the second point of view (numerical probabilities) but all the results of this paper are significant from both perspectives. Let us mention that the difficulty of the analysis and the interest of the result depend on the regularity of the test function f. It turns out that if f is a smooth function, then the analysis of the error is rather simple, using a Taylor type expansion in short time first, and a concatenation argument after. However, the study is much more subtle if f is simply a bounded and measurable test function - this is the so called convergence in total variation distance. A lot of work has been done in this direction in the case of the CLT. In particular, Bhattacharya and Rao [15] obtained the convergence when $f(x) = \mathbb{1}_A(x)$ where A is a measurable set that belongs to a large class (including convex sets). From that point, one would hope to get such results for every measurable set A and consequently for every measurable and bounded test function f. Eventually, the seminal result of Prokhorov [62] clarified this point: He proved that the convergence in total variation in the CLT may not be obtained without some regularity assumptions on the law of Z_k. Essentially, one has to assume that the law of Z_k has an absolute continuous component. In our framework this hypothesis has to be slightly strengthened. We assume that Z_k verifies the Doublin's condition (see (1.8)). In this way, we extract some regular noise and use it in order to build some integration by parts formulas (inspired from Malliavin calculus). Then, we use those formulas to regularize the test function f and finally to achieve our error analysis.

Main results

Let us now present our results with more details. In order to do it, we have to introduce some notations. For fixed $T > 0$ and $n \in \mathbb{N}^*$, we define the homogeneous time grid $\pi_{T,n} = \{t_k^n = kT/n, k \in \mathbb{N}\}$. We consider the d dimensional Markov chain

$$X^n_{t^n_{k+1}} = \psi_k(X^n_{t^n_k}, \frac{Z_{k+1}}{\sqrt{n}}, \delta^n_{k+1}), \qquad k \in \mathbb{N}, \tag{1.1}$$

where $\psi_k : \mathbb{R}^d \times \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}^d$ is a smooth function such that $\psi_k(x, 0, 0) = x$, and $Z_k \in \mathbb{R}^N$, $k \in \mathbb{N}^*$, is a sequence of independent and centered random variables and $\sup_{k\in\mathbb{N}^*} \delta_k^n \leqslant C/n$. The semigroup of the Markov chain $(X_t^n)_{t\in\pi_{T,n}}$ is denoted by $(Q_t^n)_{t\in\pi_{T,n}}$ and its transition probabilities are given by $\nu^n_{k+1}(x, dy) = \mathbb{P}(X^n_{t^n_{k+1}} \in dy | X^n_{t^n_k} = x)$, $k \in \mathbb{N}$. We recall that for $t \in \pi_{T,n}$,

$Q_t^n f(x) = \mathbb{E}[f(X_t^n)|X_0^n = x]$. We will also consider a Markov process in continuous time $(X_t)_{t\geqslant 0}$ with semigroup $(P_t)_{t\geqslant 0}$ and we define $\mu_{k+1}^n(x,dy) = \mathbb{P}(X_{t_{k+1}^n} \in dy | X_{t_k^n} = x)$.

Moreover, for $f \in \mathcal{C}^\infty(\mathbb{R}^d)$ and for a multi-index $\alpha = (\alpha_1, \cdots, \alpha_d) \in \mathbb{N}^d$ we denote $|\alpha| = \alpha_1 + ... + \alpha_d$ and $\partial_\alpha f = (\partial_1)^{\alpha_1} \ldots (\partial_d)^{\alpha_d} f = \partial_x^\alpha f(x) = \partial_{x_1}^{\alpha_1} \ldots \partial_{x_d}^{\alpha_d} f(x)$. We include the multi-index $\alpha = (0, ..., 0)$ and in this case $\partial_\alpha f = f$. We will use the norms

$$\|f\|_{q,\infty} = \sup_{x\in\mathbb{R}^d} \sum_{0\leqslant|\alpha|\leqslant q} |\partial_\alpha f(x)|, \qquad \|f\|_{q,1} = \sum_{0\leqslant|\alpha|\leqslant q} \int_{\mathbb{R}^d} |\partial_\alpha f(x)| dx.$$

In particular $\|f\|_{0,\infty} = \|f\|_\infty$ is the usual supremum norm and we will denote $\mathcal{C}_b^q(\mathbb{R}^d) = \{f \in \mathcal{C}^q(\mathbb{R}^d), \|f\|_{q,\infty} < \infty\}$ and $C_c^q(\mathbb{R}^d) \subset C^q(\mathbb{R}^d)$ the set of functions with compact support.

A first standard result is the following: Let us assume that there exists $h > 0$, $q \in \mathbb{N}$ such that for every $f \in \mathcal{C}^q(\mathbb{R}^d)$, $k \in \mathbb{N}^*$ and $x \in \mathbb{R}^d$,

$$|\mu_k^n f(x) - \nu_k^n f(x)| = |\textstyle\int f(y)\mu_k^n(x,dy) - \int f(y)\nu_k^n(x,dy)| \leqslant C\|f\|_{q,\infty}/n^{1+h}. \tag{1.2}$$

Then, for all $T \geqslant 0$, there exists $C \geqslant 1$ such that we have

$$\sup_{t\in\pi_{T,n};t\leqslant T} \|P_t f - Q_t^n f\|_\infty \leqslant C\|f\|_{q,\infty}/n^h. \tag{1.3}$$

It means that $(X_t^n)_{t\in\pi_{T,n}}$ is an approximation scheme of weak order h for the Markov process $(X_t)_{t\geqslant 0}$. In the case of the Euler scheme for diffusion processes, this result, with $h = 1$, has initially been proved in the seminal papers of Milstein [55] and of Talay and Tubaro [68] (see also [39]). Similar results were obtained in various situations: Diffusion processes with jumps (see [63], [36]) or diffusion processes with boundary conditions (see [30], [18], [31]). An overview of this subject is given in [38]. More recently, approximation schemes of higher orders (*e.g.*, $h = 2$), based on cubature methods, have been introduced and studied by Kusuoka [43], Lyons [53], Ninomiya, Victoir [57] or Alfonsi [3]. The reader may also refer to the work of Kohatsu-Higa and Tankov [40] for a higher weak order scheme for jump processes.

Another result concerns convergence in total variation distance. We want to obtain (1.3) with $\|f\|_{q,\infty}$ replaced by $\|f\|_\infty$ when f is a measurable function. In the case of the Euler scheme for diffusion processes, a first result of this type has been obtained by Bally and Talay [11], [12] using the Malliavin calculus (see also Guyon [34]). Afterwards Konakov, Menozzi and Molchanov [41], [42] obtained similar results using a parametrix method. Recently Kusuoka [44] obtained estimates of the error in total variation distance for the Victoir Ninomiya scheme (which corresponds to the case $h = 2$). We will obtain a similar result using our approach. Moreover, we give estimates of the rate of convergence of the density function and its derivatives.

Regularization properties.

We first remark that the crucial property which is used in order to replace $\|f\|_{q,\infty}$ by $\|f\|_\infty$ in (1.3), is the regularization property of the semigroup. Let us be more precise: Let $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ an increasing function, $q \in \mathbb{N}$ be fixed. Given the time grid $\pi_{T,n} = \{t_k^n = kT/n, k \in \mathbb{N}\}$, we say that a semigroup $(P_t^n)_{t\in\pi_{T,n}}$ satisfies $R_{q,\eta}$, if

$$R_{q,\eta} \qquad \forall t \in \pi_{T,n}, t > 0, \quad \|P_t^n f\|_{q,\infty} \leqslant \frac{C}{t^{\eta(q)}}\|f\|_\infty. \tag{1.4}$$

We also introduce a dual regularization property: We consider the dual semigroup $P_t^{n,*}$ (*i.e.* $\langle P_t^{n,*}g,f\rangle=\langle g,P_t^n f\rangle$ with the scalar product in $\mathrm{L}^2(\mathbb{R}^d)$) and we assume that

$$R_{q,\eta}^* \qquad \forall t\in\pi_{T,n}, t>0,\quad \|P_t^{n,*}f\|_{q,1}\leqslant\frac{C}{t^{\eta(q)}}\|f\|_1. \tag{1.5}$$

Finally, we consider the following stronger regularization property: For every multi-index α,β with $|\alpha|+|\beta|=q$,

$$\overline{R}_{q,\eta} \qquad \forall t\in\pi_{T,n}, t>0,\quad \|\partial_\alpha P_t^n\partial_\beta f\|_\infty\leqslant\frac{C}{t^{\eta(q)}}\|f\|_\infty. \tag{1.6}$$

We notice that $\overline{R}_{q,\eta}$ implies both $R_{q,\eta}$ and $R_{q,\eta}^*$ and that a semigroup satisfying $\overline{R}_{q,\eta}$ is absolutely continuous with respect to the Lebesgue measure.
In addition to (1.2), we will also suppose that the following dual estimate of the error in short time holds:

$$|\langle g,(\mu_k^n-\nu_k^n)f\rangle|\leqslant C\|g\|_{q,1}\|f\|_\infty/n^{1+h}. \tag{1.7}$$

Using those hypothesis, we can obtain a first result.

Theorem 1.1.1. *We recall that $T>0$ and $n\in\mathbb{N}^*$. We fix $h>0$, $q\in\mathbb{N}$ and we assume that the short time estimates (1.2) and (1.7) hold (with this q and h). Moreover, we assume that (1.4) holds for $(P_t)_{t\in\pi_{T,n}}$ and that (1.5) holds for $(Q_t^n)_{t\in\pi_{T,n}}$. Then, for every $S\in[T/n,T/2)$,*

$$\forall t\in\pi_{T,n}, t\geqslant 2S,\quad \|P_tf-Q_t^nf\|_\infty\leqslant\frac{C}{S^{\eta(q)}}\|f\|_\infty/n^h.$$

Integration by parts formulas.
Once we have this abstract result, the following step is to give sufficient conditions in order to obtain $R_{q,\eta}$, $R_{q,\eta}^*$ and $\overline{R}_{q,\eta}$. The method we adopt in this paper is to use Malliavin type integration by parts formulas based on the noise $Z_k\in\mathbb{R}^N$, $k\in\mathbb{N}^*$. Then we will have to bound the weights that appear in those formulas and the regularization properties will follow.

In order to obtain those estimates, we assume that the law of each Z_k is locally lower bounded by the Lebesgue measure: There exists some $z_{*,k}\in\mathbb{R}^N$ and $r_*,\varepsilon_*>0$ such that for every measurable set $A\subset B_{r_*}(z_{*,k})$ one has

$$\mathbb{P}(Z_k\in A)\geqslant\varepsilon_*\lambda(A) \tag{1.8}$$

where λ is the Lebesgue measure. If this property holds then a "splitting method" can be used in order to represent Z_k as

$$\frac{Z_k}{\sqrt{n}}=\chi_kU_k+(1-\chi_k)V_k,$$

where χ_k,U_k,V_k are independent random variables, χ_k is a Bernoulli random variable and $\sqrt{n}U_k\sim\varphi_{r_*}(u)du$ with $\varphi_{r_*}\in\mathcal{C}^\infty(\mathbb{R}^N)$. Then we use the abstract Malliavin calculus based on U_k, developed in [9] and [6], in order to obtain integration by parts formulas. The crucial point is that the density φ_{r_*} of $\sqrt{n}U_k$ is smooth and we control its logarithmic derivatives. Using this property, we build integration by parts formulas and we obtain relevant estimates for the weights which appear in these formulas. It is worth mentioning that, a variant of the Malliavin calculus based on a similar splitting method has already been used by Nourdin and Poly [59] (see also [58] and [48]). They use the so called Γ calculus introduced by Bakry, Gentil and

Ledoux [5]. Roughly speaking, the difference between our approach and the one in [5] is the following: Our construction is similar to the "simple functionals" approach in Malliavin calculus and has the derivative operator as basic object. In contrast, in the Γ calculus, the basic object is the Ornstein Uhlenbeck operator.
In order to state the main result of our paper, we introduce some additional assumptions:

$$\forall p \in \mathbb{N}, \quad \sup_{k\in\mathbb{N}^*} \mathbb{E}[|Z_k|^p] < \infty, \tag{1.9}$$

$$\forall r \in \mathbb{N}^*, \quad \sup_{k\in\mathbb{N}} \|\psi_k\|_{1,r,\infty} = \sum_{|\alpha|=0}^{r} \sum_{|\beta|+|\gamma|=1}^{r-|\alpha|} \|\partial_x^\alpha \partial_z^\beta \partial_t^\gamma \psi_k\|_\infty < \infty \tag{1.10}$$

$$\exists \lambda_* > 0,\ \forall k \in \mathbb{N}, \quad \inf_{x\in\mathbb{R}^d} \inf_{|\eta|=1} \sum_{i=1}^{N} \langle \partial_{z_i}\psi_k(x,0,0), \eta\rangle^2 \geqslant \lambda_*. \tag{1.11}$$

Moreover, we introduce the following regularized version of the approximation scheme $(X_t^n)_{t\in\pi_{T,n}}$:

$$\forall t \in \pi_{T,n}, \quad X_t^{n,\theta}(x) = \frac{1}{n^\theta} G + X_t^n(x),$$

with G a standard normal random variable independent from $X_{t_k^n}^n$ and $\theta \geqslant h+1$. Here $X_t^n(x)$ is the Markov chain which starts from x: $X_0^n(x) = x$. We denote

$$Q_t^{n,\theta}(x,dy) = \mathbb{P}(X_t^{n,\theta}(x) \in dy) = p_t^{n,\theta}(x,y)dy.$$

Theorem 1.1.2. *We recall that $T > 0$ and $n \in \mathbb{N}^*$. We fix $h > 0$, $q \in \mathbb{N}$ and and we consider a Markov semigroup $(P_t)_{t\geqslant 0}$ and the discrete Markov chain $(Q_t^n)_{t\in\pi_{T,n}}$ defined in (1.1). We assume that the short time estimates (1.2) and (1.7) hold (with this q and h). Moreover, we assume (1.8), (1.9), (1.10) and (1.11).*

A. *For every $S \in [T/n, T/2)$, we have*

$$\forall t \in \pi_{T,n}, t \in (2S,T], \quad \|P_t f - Q_t^n f\|_\infty \leqslant \frac{C}{(\lambda_* S)^{\eta(q)}} \|f\|_\infty / n^h. \tag{1.12}$$

B. *For every $t > 0$, $P_t(x,dy) = p_t(x,y)dy$ with $(x,y) \mapsto p_t(x,y)$ belonging to $\mathcal{C}^\infty(\mathbb{R}^d \times \mathbb{R}^d)$.*

C. *For every $x_0, y_0, R > 0, \varepsilon \in (0,1)$ and every multi-index α, β, we have*

$$\forall t \in \pi_{T,n}, t \in (2S,T], \quad \sup_{\overline{B}_R(x_0,y_0)} |\partial_x^\alpha \partial_y^\beta p_t(x,y) - \partial_x^\alpha \partial_y^\beta p_t^{n,\theta}(x,y)| \leqslant C_\varepsilon / n^{h(1-\varepsilon)}, \tag{1.13}$$

with a constant C_ε which depends on $R, x_0, y_0, S, \lambda_*, T, \varepsilon, \eta$ and on $|\alpha| + |\beta|$ (and may go to infinity as ε tends to 0). Moreover we denote $\overline{B}_R(x_0,y_0) = \{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d, |(x,y) - (x_0,y_0)| \leqslant R\}$.

We notice that (1.12) gives the total variation convergence between the semigroups $(P_t)_{t\geqslant 0}$ and $(Q_t^n)_{t\in\pi_{T,n}}$. Once the appropriate regularization properties are obtained (using the abstract Malliavin calculus), the proof of (1.12) is rather elementary. In contrast, the estimate (1.13) is based on a non trivial interpolation result recently obtained in [8]. Notice, however, that the estimate (1.13) is sub-optimal (because of $\varepsilon > 0$). We will illustrate (1.12) by taking X^n to be the Ninomiya Victoir scheme of a diffusion process. This is a variant of the result already obtained by Kusuoka [44] in the case where Z_k has a Gaussian distribution (and so the standard Malliavin calculus is available). As we have mentioned in the beginning of this paper, the random variables Z_k, $k \in \mathbb{N}^*$ have an arbitrary distribution (except the property (1.8)) and our result can be seen as an invariance principle as well.

The paper is organized as follows. In Section 1.2, we prove Theorem 1.1.1. In Section 1.3, we settle the abstract Malliavin calculus based on the splitting method. We use it in Section 1.4 in order to prove the regularization properties for the approximation scheme X^n (in fact for the regularization $X^{n,\theta}$) and we prove Theorem 1.1.2. Finally, in Section 1.5, we use the previous results in order to give estimates of the total variation distance for the Ninomiya Victoir approximation scheme. Section 1.6 is a short presentaton of the simple functionals approach in Malliavin calculus which happens to be the method that inspired the differential calculus used in this paper.

1.2 The distance between two Markov semigroups

Throughout this section the following notations will prevail. We fix $T > 0$ and we denote $n \in \mathbb{N}^*$, the number of time step between 0 and T. Then, for $k \in \mathbb{N}$ we define $t_k^n = kT/n$ and we introduce the homogeneous time grid $\pi_{T,n} = \{t_k^n = kT/n, k \in \mathbb{N}\}$ and its bounded version $\pi_{T,n}^{\tilde{T}} = \{t \in \pi_{T,n}, t \leqslant \tilde{T}\}$ for $\tilde{T} \geqslant 0$. Finally, for $S \in [0, \tilde{T})$ we will denote $\pi_{T,n}^{S,\tilde{T}} = \{t \in \pi_{T,n}^{\tilde{T}}, t > S\}$. Notice that, all the results from this paper remain true with non homogeneous time step but, for sake of simplicity, we will not consider this case. First, we state some results for smooth test functions.

1.2.1 Regular test functions

We consider a sequence of finite transition measures $\mu_k^n(x, dy)$, $k \in \mathbb{N}^*$ from $\mathbb{R}^d$ to $\mathbb{R}^d$. This means that for each fixed x and k, $\mu_k^n(x, dy)$ is a finite measure on $\mathbb{R}^d$ with the borelian σ field and for each bounded measurable function $f : \mathbb{R}^d \to \mathbb{R}$, the application

$$x \mapsto \mu_k^n f(x) := \int_{\mathbb{R}^d} f(y) \mu_k^n(x, dy)$$

is measurable. We also denote

$$|\mu_k^n| := \sup_{x \in \mathbb{R}^d} \sup_{\|f\|_\infty \leqslant 1} \Big| \int_{\mathbb{R}^d} f(y) \mu_k^n(x, dy) \Big|,$$

and, we assume that all the sequences of measures we consider in this paper satisfy:

$$\sup_{k \in \mathbb{N}^*} |\mu_k^n| < \infty. \tag{1.14}$$

Although the main application concerns the case where $\mu_k^n(x, dy)$ is a probability measure, we do not assume this here. Indeed, $\mu_k^n(x, dy)$ is only supposed be a signed measure of finite (but arbitrary) total mass. This is because one may use the results from this section not only in order to estimate the distance between two semigroups but also in order to obtain an expansion of the error.

Now we associate the sequence of measures μ^n to the time grid $\pi_{T,n}$.

Definition 1.2.1. *We define the discrete semigroup P^n in the following way.*

$$P_0^n f(x) = f(x), \qquad P^n_{t^n_{k+1}} f(x) = P^n_{t^n_k}\mu^n_{k+1} f(x) = P^n_{t^n_k}\int_{\mathbb{R}^d} f(y)\mu^n_{k+1}(x,dy).$$

More generally, we define $(P_{t,s})_{t,s\in\pi_{T,n};t\leqslant s}$ by

$$P^n_{t^n_k,t^n_k} f(x) = f(x), \qquad \forall k,r\in\mathbb{N}^*, k\leqslant r\ P^n_{t^n_k,t^n_{r+1}} f(x) = P^n_{t^n_k,t^n_r}\mu^n_{r+1} f(x).$$

We notice that for $t,s,u\in\pi_{T,n}$, $t\leqslant s\leqslant u$, we have the semigroup property $P^n_{t,u}f = P^n_{t,s}P^n_{s,u}f$. We will consider the following hypothesis: Let $q\in\mathbb{N}$ and $t\leqslant s\in\pi_{T,n}$. If $f\in\mathcal{C}^q(\mathbb{R}^d)$ then $P_{t,s}f\in\mathcal{C}^q(\mathbb{R}^d)$ and

$$\sup_{t,s\in\pi_{T,n};t\leqslant s}\|P^n_{t,s}f\|_{q,\infty}\leqslant C\|f\|_{q,\infty}. \tag{1.15}$$

Notice that (1.14) implies that (1.15) holds for $q=0$.

We consider now a second sequence of finite transition measures $\nu^n_k(x,dy), k\in\mathbb{N}^*$. Moreover, we introduce the corresponding semigroup Q^n defined in a similar way as P^n with μ^n replaced by ν^n which also satisfies (1.14) and (1.15).

We aim to estimate the distance between P^nf and Q^nf in terms of the distance between the transition measures $\mu^n_k(x,dy)$ and $\nu^n_k(x,dy)$, so we denote

$$\Delta^n_k = \mu^n_k - \nu^n_k.$$

$(P^n_t)_{t\in\pi_{T,n}}$ can be seen as a semigroup in continuous time, $(P_t)_{t\geqslant 0}$, considered on the time grid $\pi_{T,n}$, while $(Q_t)_{t\in\pi_{T,n}}$ would be its approximation discrete semigroup. Let $q\in\mathbb{N}$, $h\geqslant 0$ be fixed. We introduce a short time error approximation assumption: There exists a constant $C>0$ (depending on q only) such that for every $k\in\mathbb{N}^*$, we have

$$E_n(h,q) \qquad \|\Delta^n_k f\|_\infty\leqslant C\|f\|_{q,\infty}/n^{h+1}. \tag{1.16}$$

Proposition 1.2.1. *Let $q\in\mathbb{N}$ be fixed. Suppose that ν^n satisfies (1.15) for this q and that we have $E_n(h,q)$ (see (1.16)). Then for every $f\in\mathcal{C}^q(\mathbb{R}^d)$,*

$$\sup_{t\in\pi^T_{T,n}}\|P^n_t f - Q^n_t f\|_\infty\leqslant C\|f\|_{q,\infty}/n^h. \tag{1.17}$$

Proof. Let $m\in\mathbb{N}^*$, $m\leqslant n$. We have

$$\begin{aligned}\|P^n_{t^n_m}f - Q^n_{t^n_m}f\|_\infty &\leqslant \sum_{k=0}^{m-1}\|P^n_{t^n_k}P^n_{t^n_k,t^n_{k+1}}Q^n_{t^n_{k+1},t^n_m}f - P^n_{t^n_k}Q^n_{t^n_k,t^n_{k+1}}Q^n_{t^n_{k+1},t^n_m}f\|_\infty \\ &= \sum_{k=0}^{m-1}\|P^n_{t^n_k}\Delta^n_{k+1}Q^n_{t^n_{k+1},t^n_m}f\|_\infty.\end{aligned} \tag{1.18}$$

Using (1.14) for μ^n, (1.16) and then (1.15) for ν^n, we obtain

$$\|P^n_{t^n_{k+1},t^n_m}\Delta^n_{k+1}Q^n_{t^n_k}f\|_\infty\leqslant C\|\Delta^n_{k+1}Q^n_{t^n_k}f\|_\infty\leqslant C\|Q^n_{t^n_k}f\|_{q,\infty}/n^{h+1}\leqslant C\|f\|_{q,\infty}/n^{1+h}.$$

Summing over $k=0,...,m-1$, we conclude. $\square$

1.2.2 Measurable test functions (convergence in total variation distance)

The estimate (1.17) requires a lot of regularity for the test function f. We aim to show that, if the semigroups at work have a regularization property, then we may obtain estimates of the error for measurable and bounded test functions. In order to state this result we have to give some hypothesis on the adjoint semigroup. Let $q \in \mathbb{N}$. We assume that there exists a constant $C \geqslant 1$ such that for every measurable and bounded function f and any $g \in \mathcal{C}^q(\mathbb{R}^d)$

$$E_n^*(h,q) \qquad |\langle g, \Delta_k^n f\rangle| \leqslant C\|g\|_{q,1}\|f\|_\infty / n^{1+h}. \tag{1.19}$$

where $\langle g, f\rangle = \int g(x)f(x)dx$ is the scalar product in $\mathrm{L}^2(\mathbb{R}^d)$.
Our regularization hypothesis is the following. Let $q \in \mathbb{N},\ S>0$ and $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ an increasing function be given. We assume that there exists a constant $C \geqslant 1$ such that

$$R_{q,\eta}(S) \qquad \forall t,s \in \pi_{T,n},\ \text{with } S \leqslant s-t, \quad \|P_{t,s}^n f\|_{q,\infty} \leqslant \frac{C}{S^{\eta(q)}}\|f\|_\infty. \tag{1.20}$$

We also consider the "adjoint regularization hypothesis". We assume that there exists an adjoint semigroup $P_{t,s}^{n,*}$, that is

$$\left\langle P_{t,s}^{n,*} g, f\right\rangle = \left\langle g, P_{t,s}^n f\right\rangle$$

for every measurable and bounded function f and every function $g \in \mathcal{C}_c^\infty(\mathbb{R}^d)$. We assume that $P_{t,s}^{n,*}$ satisfies

$$R_{q,\eta}^*(S) \qquad \forall t,s \in \pi_{T,n},\ \text{with } S \leqslant s-t, \quad \|P_{t,s}^{n,*} f\|_{q,1} \leqslant \frac{C}{S^{\eta(q)}}\|f\|_1. \tag{1.21}$$

Notice that a sufficient condition in order that $R_{q,\eta}^*(S)$ holds is the following: For every multi index α with $|\alpha| \leqslant q$

$$\forall t,s \in \pi_{T,n},\ \text{with } S \leqslant s-t, \quad \|P_{t,s}^n \partial_\alpha f\|_\infty \leqslant \frac{C}{S^{\eta(q)}}\|f\|_\infty. \tag{1.22}$$

Indeed:

$$\begin{aligned}
\|\partial_\alpha P_{t,s}^{n,*} f\|_1 &\leqslant \sup_{\|g\|_\infty \leqslant 1} |\left\langle \partial_\alpha P_{t,s}^{n,*} f, g\right\rangle| = \sup_{\|g\|_\infty \leqslant 1} |\left\langle f, P_{t,s}^n(\partial_\alpha g)\right\rangle| \\
&\leqslant \|f\|_1 \sup_{\|g\|_\infty \leqslant 1} \|P_{t,s}^n(\partial_\alpha g)\|_\infty \leqslant \frac{C}{S^{\eta(q)}}\|f\|_1.
\end{aligned}$$

Proposition 1.2.2. *Let $q \in \mathbb{N},\ \ h \geqslant 0,\ S \in [T/n, T/2)$ and $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ an increasing function be fixed. We assume that $E_n(h,q)$ (see (1.16)) and $E_n^*(h,q)$ (see (1.19)) hold for P^n and Q^n. We also suppose that P^n satisfies $R_{q,\eta}(S)$ (see (1.20)) and Q^n satisfies $R_{q,\eta}^*(S)$ (see (1.21)) and that (1.15) hold with $q=0$ for both of them. Then,*

$$\sup_{t \in \pi_{T,n}^{2S,T}} \|P_t^n f - Q_t^n f\|_\infty \leqslant \frac{C}{S^{\eta(q)}}\|f\|_\infty / n^h.$$

Proof. Using a density argument we may assume that $f \in \mathcal{C}(\mathbb{R}^d)$. Moreover, by (1.18), it is sufficient to prove that

$$\|Q_{t_k^n}^n \Delta_{k+1}^n P_{t_{k+1}^n, t_m^n}^n f\|_\infty \leqslant \frac{C}{S^{\eta(q)}}\|f\|_\infty / n^{1+h},$$

for $m \in \{2, \ldots, n\}$. Since $t_m^n > 2S$ we have $t_k^n \geqslant S$ or $t_m^n - t_{k+1}^n \geqslant S$. Suppose first that $t_m^n - t_{k+1}^n \geqslant S$. Using (1.14) for Q^n, (1.16) and (1.20) for P^n,

$$\|Q_{t_k^n}^n \Delta_{k+1}^n P_{t_{k+1}^n, t_m^n}^n f\|_\infty \leqslant C \|\Delta_{k+1}^n P_{t_{k+1}^n, t_m^n}^n f\|_\infty \leqslant C \|P_{t_{k+1}^n, t_m^n}^n f\|_{q,\infty}/n^{1+h} \leqslant CS^{-\eta(q)} \|f\|_\infty / n^{1+h}.$$

Suppose now that $t_k^n \geqslant S$. We take $\phi_\varepsilon(x) = \varepsilon^{-d}\phi(\varepsilon^{-1}x)$ with $\phi \in \mathcal{C}_c(\mathbb{R}^d)$, $\phi \geqslant 0$. Then, for a fixed x_0, we define $\phi_{\varepsilon,x_0}(x) = \phi_\varepsilon(x - x_0)$. By (1.15), $Q_{t_k^n}^n \Delta_{k+1}^n P_{t_{k+1}^n, t_m^n}^n f$ is continuous. Then

$$|Q_{t_k^n}^n \Delta_{k+1}^n P_{t_{k+1}^n, t_m^n}^n f(x_0)| = \lim_{\varepsilon \to 0} |\langle \phi_{\varepsilon,x_0}, Q_{t_k^n}^n \Delta_{k+1}^n P_{t_{k+1}^n, t_m^n}^n f \rangle|.$$

Using (1.19), (1.21) for Q^n and then (1.14) for P^n, we obtain

$$\begin{aligned}|\langle \phi_{\varepsilon,x_0}, Q_{t_k^n}^n \Delta_{k+1}^n P_{t_{k+1}^n, t_m^n}^n f \rangle| &= |\langle Q_{t_k^n}^{n,*} \phi_{\varepsilon,x_0}, \Delta_{k+1}^n P_{t_{k+1}^n, t_m^n}^n f \rangle| \leqslant C \|Q_{t_k^n}^{n,*} \phi_{\varepsilon,x_0}\|_{q,1} \|P_{t_{k+1}^n, t_m^n}^n f\|_\infty / n^{1+h} \\ &\leqslant CS^{-\eta(q)} \|\phi_{\varepsilon,x_0}\|_1 \|f\|_\infty / n^{1+h}\end{aligned}$$

and since $\|\phi_{\varepsilon,x_0}\|_1 = \|\phi\|_1 \leqslant C$, the proof is completed. $\square$

In concrete applications the following slightly more general variant of the above proposition will be useful.

Proposition 1.2.3. *Let $q \in \mathbb{N}$, $h \geqslant 0$, $S \in [T/n, T/2)$ and $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ an increasing function be fixed. We assume that $E_n(h,q)$ (see (1.16)) and $E_n^*(h,q)$ (see (1.19)) hold for P^n and Q^n. Moreover, we assume that there exists some kernels $(\overline{P}_{t,s}^n)_{t,s \in \pi_{T,n}; t \leqslant s}$ which satisfies $R_{q,\eta}(S)$ (see(1.20)) and $(\overline{Q}_{t,s}^n)_{t,s \in \pi_{T,n}; t \leqslant s}$ which satisfies $R_{q,\eta}^*(S)$ (see (1.21)) and that (1.15) hold with $q = 0$ for both of them. We also assume that for every $t, s \in \pi_{T,n}$ with $s - t \geqslant S$,*

$$\|Q_{t,s}^n f - \overline{Q}_{t,s}^n f\|_\infty + \|P_{t,s}^n f - \overline{P}_{t,s}^n f\|_\infty \leqslant CS^{-\eta(q)} \|f\|_\infty / n^{h+1}. \tag{1.23}$$

Then,

$$\sup_{t \in \pi_{T,n}^{2S,T}} \|P_t^n f - Q_t^n f\|_\infty \leqslant C \sup_{k \leqslant n} (|\mu_k^n| + |\nu_k^n|) S^{-\eta(q)} \|f\|_\infty / n^h.$$

Remark 1.2.1. *Notice that $\overline{P}^n$ and $\overline{Q}^n$ are not supposed to satisfy the semigroup property and are not directly related to μ^n and ν^n.*

Proof. The proof follows the same line as the one of the previous proposition. Suppose first that $t_m^n - t_k^n \geqslant S$. Then, (1.14) implies

$$\begin{aligned}\|Q_{t_k^n}^n \Delta_{k+1}^n P_{t_{k+1}^n, t_m^n}^n f\|_\infty &\leqslant \|Q_{t_k^n}^n \Delta_{k+1}^n \overline{P}_{t_{k+1}^n, t_m^n}^n f\|_\infty + \|Q_{t_k^n}^n \Delta_{k+1}^n (P_{t_{k+1}^n, t_m^n}^n - \overline{P}_{t_{k+1}^n, t_m^n}^n) f\|_\infty \\ &\leqslant \|\Delta_{k+1}^n \overline{P}_{t_{k+1}^n, t_m^n} f\|_\infty + \|\Delta_{k+1}^n (P_{t_{k+1}^n, t_m^n}^n - \overline{P}_{t_{k+1}^n, t_m^n}^n) f\|_\infty.\end{aligned}$$

Since $\overline{P}^n$ verifies $R_{q,\eta}(S)$, we deduce from (1.16) that

$$\|\Delta_{k+1}^n \overline{P}_{t_k^n}^n f\|_\infty \leqslant C \|\overline{P}_{t_k^n}^n f\|_{q,\infty} / n^{h+1} \leqslant CS^{-\eta(q)} \|f\|_\infty / n^{h+1}.$$

Using (1.23), it follows

$$\begin{aligned}|\Delta_{k+1}^n (P_{t_{k+1}^n, t_m^n}^n - \overline{P}_{t_{k+1}^n, t_m^n}^n) f(x)| &\leqslant |\int (P_{t_{k+1}^n, t_m^n}^n - \overline{P}_{t_{k+1}^n, t_m^n}^n) f(y) \nu_{k+1}(x, dy)| \\ &\quad + |\int (P_{t_{k+1}^n, t_m^n}^n - \overline{P}_{t_{k+1}^n, t_m^n}^n) f(y) \mu_{k+1}(x, dy)| \\ &\leqslant (|\nu_{k+1}^n| + |\mu_{k+1}^n|) \|(P_{t_{k+1}^n, t_m^n}^n - \overline{P}_{t_{k+1}^n, t_m^n}^n) f\|_\infty \\ &\leqslant C(|\nu_{k+1}^n| + |\mu_{k+1}^n|) S^{-\eta(q)} \|f\|_\infty / n^{h+1}.\end{aligned}$$

Suppose now that $t_k^n \geqslant S$. We write

$$\|Q^n_{t^n_k}\Delta^n_{k+1}P^n_{t^n_{k+1},t^n_m}f\|_\infty \leqslant \|\overline{Q}^n_{t^n_k}\Delta^n_{k+1}P^n_{t^n_{k+1},t^n_m}f\|_\infty + \|(Q^n_{t^n_k}-\overline{Q}^n_{t^n_k})\Delta^n_{k+1}P^n_{t^n_k,t^n_m}f\|_\infty.$$

In order to bound $\|\overline{Q}^n_{t^n_k}\Delta^n_{k+1}P^n_{t^n_{k+1},t^n_m}f\|_\infty$ we use the same reasoning as in the proof of the previous proposition. And the second term is bounded using (1.23). □

1.2.3 Convergence of the density functions

In this section we will consider a Markov semigroup $(P_t)_{t\geqslant 0}$ and we will give an approximation result and a regularity criterion for it. The regularization property that we assume for the approximation processes is stronger than the one considered in the previous section and, instead of Proposition 1.2.2 we will use a general approximation result based on an interpolation inequality, proved in [8]. We recall that we have fixed $T>0$ and that, and for $n\in\mathbb{N}^*$ we denote $t_k^n = kT/n$. For $k\in\mathbb{N}^*$, we consider $\mu_k^n(x,dy)=\mu^n(x,dy)=P_{T/n}(x,dy)$, for all $k\in\mathbb{N}$, the homogeneous sequence of finite transition measures which satisfy (1.15). To this sequence of measures, we associate the discrete version $(P_t^n)_{t\in\pi_{T,n}}$ of P such that for all $t,s\in\pi_{T,n}$, $t\leqslant s$, $P^n_{t,s}f(x)=P_{s-t}f(x)$. Moreover we introduce a sequence of transition probability measures $\nu_k^n(x,dy)$, $k\in\mathbb{N}^*$, and the corresponding discrete semigroups $Q^n(x,dy)$ defined by $Q^n_{t,t}=Id$ and $Q^n_{t^n_k,t^n_{r+1}} = Q^n_{t^n_k,t^n_r}\nu^n_{r+1}$. We recall that for all $t\in\pi_{T,n}$ then $Q^n_t f = Q^n_{0,t}f$. We assume that for $f\in\mathcal{C}^q(\mathbb{R}^d)$, we have $Q^n_{t,s}f\in\mathcal{C}^q(\mathbb{R}^d)$ for all $t,s\in\pi_{T,n}, t\leqslant s$, and it verifies (1.15) :

$$\sup_{t,s\in\pi_{T,n};t\leqslant s}\|Q^n_{t,s}f\|_{q,\infty}\leqslant C\|f\|_{q,\infty}.$$

For $h>0$ and $q\in\mathbb{N}$, we assume that we have (1.16) and (1.19):

$$E_n(h,q)\qquad \|(\mu^n-\nu^n_k)f\|_\infty\leqslant C\|f\|_{q,\infty}/n^{1+h}.$$

and,

$$E^*_n(h,q)\qquad |\langle g,(\mu^n-\nu^n_k)f\rangle|\leqslant C\|g\|_{q,1}\|f\|_\infty/n^{1+h}.$$

In concrete applications, it may be cumbersome to prove the regularization properties of the underlying semigroups P^n and Q^n. In order to treat this problem, we introduce now $(\overline{Q}^n_t)_{t\in\pi_{T,n}}$, a modification of $(Q^n_t)_{t\in\pi_{T,n}}$ in the sense that for every measurable and bounded function $f:\mathbb{R}^d\to\mathbb{R}$, we have

$$\forall t,s\in\pi_{T,n},\ \text{with}\ S\leqslant s-t,\quad \|Q^n_{t,s}f-\overline{Q}^n_{t,s}f\|_\infty\leqslant\frac{C}{S^{\eta(q)}}\|f\|_\infty/n^{h+1}. \tag{1.24}$$

We assume that $(\overline{Q}^n_t)_{t\in\pi_{T,n}}$ satisfies the following strong regularization property. We fix $q\in\mathbb{N}$ $S,\eta>0$, and we assume that for every multi-index α,β with $|\alpha|+|\beta|\leqslant q$ and $f\in\mathcal{C}^q(\mathbb{R}^d)$ one has

$$\overline{R}_{q,\eta}(S)\qquad \forall t,s\in\pi_{T,n},\ \text{with}\ S\leqslant s-t,\quad \|\partial_\alpha\overline{Q}^n_{t,s}\partial_\beta f\|_\infty\leqslant CS^{-\eta(q)}\|f\|_\infty. \tag{1.25}$$

Notice that if $\overline{R}_{q+2d,\eta}(S)$ holds, then for all $t\in\pi_{T,n}$, there exists $\overline{p}^n_t\in\mathcal{C}^q(\mathbb{R}^d\times\mathbb{R}^d)$ such that $\overline{Q}^n_t(x,dy)=\overline{p}^n_t(x,y)dy$. Moreover, if $t\geqslant S$, then for every $|\alpha|+|\beta|\leqslant q$, we have

$$\sup_{(x,y)\in\mathbb{R}^d\times\mathbb{R}^d}|\partial^\alpha_x\partial^\beta_y\overline{p}^n_t(x,y)|\leqslant CS^{-\eta(q+2d)}. \tag{1.26}$$

Indeed, let $f_\zeta : \mathbb{R}^d \to \mathbb{C}, x \mapsto e^{-i\langle \zeta, x\rangle}$. Using the Fourier representation of the density function, we have

$$\overline{p}_t^n(x,y) = \int_{\mathbb{R}^d} e^{i\langle \zeta, y\rangle} \overline{Q}_t^n f_\zeta(x) d\zeta$$

Now we notice that $\partial_y^\beta f_\zeta(y) = f_\zeta(y)(-i)^{|\beta|} \prod_{i=1}^{|\beta|} \zeta_{\beta_i}$ and it follows that for all $x, y, \in \mathbb{R}^d$,

$$\begin{aligned}
\partial_x^\alpha \partial_y^\beta \overline{p}_t^n(x,y) &= \int_{\mathbb{R}^d} i^{|\beta|} (\prod_{i=1}^{|\beta|} \zeta_{\beta_i}) e^{i\langle \zeta, y\rangle} \partial_x^\alpha (\overline{Q}_t^n f_\zeta)(x) d\zeta \\
&= \int_{[-1,1]^d} i^{|\beta|} (\prod_{i=1}^{|\beta|} \zeta_{\beta_i}) e^{i\langle \zeta, y\rangle} \partial_x^\alpha (\overline{Q}_t^n f_\zeta)(x) d\zeta + \int_{\mathbb{R}^d \setminus [-1,1]^d} i^{|\beta|} (\prod_{i=1}^{|\beta|} \zeta_{\beta_i}) e^{i\langle \zeta, y\rangle} \partial_x^\alpha (\overline{Q}_t^n f_\zeta)(x) d\zeta \\
&=: I + J
\end{aligned}$$

Since $\|f_\zeta\|_\infty = 1$, we use (1.25) and we obtain: $|I| \leqslant CS^{-\eta(|\alpha|)} \leqslant CS^{-\eta(q)}$. Moreover, for any multi-index β', we have

$$J = (-1)^{|\beta|} i^{|\beta'|} \int_{\mathbb{R}^d \setminus [-1,1]^d} \frac{e^{i\langle \zeta, y\rangle}}{\prod_{i=1}^{|\beta'|} \zeta_{\beta'_i}} \partial_x^\alpha (\overline{Q}_t^n \partial_{\beta'} \partial_\beta f_\zeta)(x) d\zeta.$$

We take $\beta' = (2, \ldots, 2)$ and we obtain similarly $|J| \leqslant CS^{-\eta(q+2d)}$. We gather all the terms together and we obtain Equation (1.26). Finally, we recall that the regularization properties $R_{q,\eta}(S)$ and $R^*_{q,\eta}(S)$ hold when $\overline{R}_{q,\eta}(S)$ is satisfied.

Theorem 1.2.1. *We recall that $T > 0$ and $n \in \mathbb{N}^*$. We have the following properties.*

A. *We fix $q \in \mathbb{N}$, h, $S \in [T/n, T/2)$ and $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ an increasing function. We assume that for every $m \in \mathbb{N}$, $m \geqslant n$, there exists some modifications $(\overline{Q}_t^m)_{t \in \pi_{T,m}}$ of $(Q_t^m)_{t \in \pi_{T,m}}$ such that (1.24) and (1.25) hold for these q, h, η and S. Moreover we assume that $E_m(h,q)$ (see (1.16)) and $E^*_m(h,q)$ (see (1.19)) hold between $(P_t^m)_{t \in \pi_{T,m}} = (P_t)_{t \in \pi_{T,m}}$ and $(Q_t^m)_{t \in \pi_{T,m}}$ and that (1.15) hold for Q^m. Then, we have*

$$\sup_{t \in \pi_{T,n}^{2S,T}} \|P_t f - Q_t^n f\|_\infty \leqslant CS^{-\eta(q)} \|f\|_\infty / n^h. \tag{1.27}$$

B. *Moreover, we suppose that the modifications $\overline{Q}$ of Q satisfy also $\overline{R}_{\overline{q},\eta}(S)$ (see (1.25)) for every $\overline{q} \in \mathbb{N}$. Then, for every $t > 0$, $P_t(x, dy) = p_t(x,y)dy$ with $(x,y) \mapsto p_t(x,y)$ belonging to $\mathcal{C}^\infty(\mathbb{R}^d \times \mathbb{R}^d)$.*

C. *For every $R > 0, \varepsilon \in (0,1)$ and every multi-index α, β with $|\alpha| + |\beta| = u$, we also have*

$$\sup_{t \in \pi_{T,n}^{2S,T}} \sup_{(x,y) \in \overline{B}_R(x_0,y_0)} |\partial_x^\alpha \partial_y^\beta p_t(x,y) - \partial_x^\alpha \partial_y^\beta \overline{p}_t^n(x,y)| \leqslant CS^{-\eta(p_{u,\varepsilon} \vee q)} / n^{h(1-\varepsilon)} \tag{1.28}$$

with a constant C which depends on R, x_0, y_0, T and on $|\alpha| + |\beta|$ and $p_{u,\varepsilon} = (u + 2d + 1 + 2\lceil (1-\varepsilon)(u+d)/(2\varepsilon) \rceil)$.

Remark 1.2.2. *The inequality (1.27) is essentially a consequence of Proposition 1.2.3. However, we may not use directly this result, because we do not assume that the semigroup $(P_t)_{t \geqslant 0}$ has the regularization property (1.20) or even the less restrictive hypothesis (1.15). It simply* *satisfies (1.14). This is a result of main interest since we have to check the regularization properties for the approximation scheme Q^n only (more precisely for every $Q^m, m \geqslant n$). Indeed, in concrete applications, it can be cumbersome to study the regularization property for P. Using this result, it is not necessary anymore. Consequently in this paper, we will only study the regularization properties of the approximation Markov chain (1.1) and we will give sufficient conditions in order to obtain those properties.*

Remark 1.2.3. *The estimate (1.28) is sub-optimal because of $\varepsilon > 0$. One may wonder if optimal estimates (with n^h instead of $n^{h(1-\varepsilon)}$) may be obtained - as it was the case in the paper of Bally and Talay [11] concerning the Euler scheme. Notice that, in the above paper, specific properties related to the dynamics of the diffusion process which gives the semigoup are used, and in particular properties of the tangent flow. For example, if $X_t(x)$ denotes the diffusion process starting from x then we have $\mathbb{E}[f'(X_t(x))] = \partial_x \mathbb{E}[f(X_t(x))(\partial_x X_t(x))^{-1}] - \mathbb{E}[f(X_t(x))\partial_x(\partial_x X_t(x))^{-1})]$. Such properties are crucial in the above paper - but are difficult to express in terms of general semigroups.*

Proof. We prove **A** first. We fix $n \in \mathbb{N}^*$. Now we introduce the sequence of discrete semigroups $((Q_t^{n,m})_{t \in \pi_{T,n}})_{m \in \mathbb{N}^*}$ defined in the following way: For all $t \in \pi_{T,n}$ we have $Q_t^{n,m} f(x) = Q_t^{nm} f(x)$. Let $m' \geqslant m$, then

$$\begin{aligned}\|Q^{n,m}_{t^n_k,t^n_{k+1}} f - Q^{n,m'}_{t^n_k,t^n_{k+1}} f\|_\infty &= \|Q^{n,m}_{t^{nm}_{mk},t^{nm}_{m(k+1)}} f - Q^{n,m'}_{t^{nm'}_{m'k},t^{nm'}_{m'(k+1)}} f\|_\infty \\ &\leqslant \|Q^{nm}_{t^{nm}_{mk},t^{nm}_{m(k+1)}} f - P^{nm}_{t^{nm}_{mk},t^{nm}_{m(k+1)}} f\|_\infty + \|P^{nm'}_{t^{nm'}_{m'k},t^{nm'}_{m'(k+1)}} f - Q^{nm'}_{t^{nm'}_{m'k},t^{nm'}_{m'(k+1)}} f\|_\infty\end{aligned}$$

Since Q^{nm} and $Q^{nm'}$ verify respectively $E_{nm}(h,q)$ and $E_{nm'}(h,q)$ and both Q^{nm} and $Q^{nm'}$ satisfy (1.15), we use the Lindeberg decomposition (1.18) in order to obtain: $\|Q^{n,m}_{t^n_k,t^n_{k+1}} f - Q^{n,m'}_{t^n_k,t^n_{k+1}} f\|_\infty \leqslant C\|f\|_{q,\infty}/(n^{h+1}m^h)$. In the same way we obtain $|\langle g, Q^{n,m}_{t^n_k,t^n_{k+1}} f - Q^{n,m'}_{t^n_k,t^n_{k+1}} f\rangle| \leqslant C\|g\|_{1,q}\|f\|_\infty/(n^{h+1}m^h)$. Now, since both Q^{nm} and $Q^{nm'}$ have modifications which satisfy (1.24) and (1.25), we use the same reasoning as in the proof of Proposition 1.2.3 and it follows that: $\forall t \in \pi^{2S,T}_{T,n}, \|Q_t^{n,m} f - Q_t^{n,m'} f\|_\infty \leqslant CS^{-\eta(q)}\|f\|_\infty/(n^h m^h)$. The sequence $((Q_t^{n,m})_{t \in \pi_{T,n}})_{m \in \mathbb{N}^*}$ is thus Cauchy and it converges toward $(P_t^n)_{t \in \pi_{T,n}}$ for smooth test functions using Proposition 1.2.1. In particular, taking $m = 1$ and letting m' tend to infinity in the previous inequality we have

$$\forall t \in \pi^{2S,T}_{T,n}, \quad \|Q_t^{n,1} f - P_t^n f\|_\infty \leqslant CS^{-\eta(q)}\|f\|_\infty/n^h,$$

which is (1.27). Let us prove **C**. We are going to use a result from [8]. First, we introduce some notations. For $q \in \mathbb{N}$, we introduce the distance d_q defined by

$$d_q(\mu,\nu) = \sup\{|\textstyle\int f d\mu - \int f d\nu\| : \|f\|_{q,\infty} \leqslant 1\}.$$

For $q,\ l \in \mathbb{N}$, $r > 1$ and $f \in \mathcal{C}^q(\mathbb{R}^d \times \mathbb{R}^d)$, we denote

$$\|f\|_{q,l,r} = \sum_{0 \leqslant |\alpha| \leqslant q} (\textstyle\int\int (1 + |x|^l + |y|^l)|\partial_\alpha f(x,y)|^r dx dy)^{1/r}.$$

Since we want to show how the constant depends from S in the right hand side of (1.28), we will use a variant of Theorem 2.11 from [8].

Proposition 1.2.4. *Let $p, \tilde{p} \in \mathbb{N}$, $m \in \mathbb{N}^*$ and $r > 1$ be given and let r^* be the conjugate of r. We consider some measures $\mu(dx, dy)$ and $\mu_{g_n}(dx, dy) = g_n(x, y)dxdy$ with $g_n \in \mathcal{C}^{p+2m}(\mathbb{R}^d \times \mathbb{R}^d)$ and we assume that there exists $K_\mu, K_{g,p,m} \geqslant 1$, $h \in \mathbb{N}^*$, such that*

$$d_{\tilde{p}}(\mu, \mu_{g_n}) \leqslant K_\mu / n^h, \quad \|g_n\|_{p+2m,2m,r} \leqslant K_{g,p,m}, \qquad \forall n \in \mathbb{N}. \tag{1.29}$$

Then $\mu(dx, dy) = g(x, y)dxdy$ where g belongs to the Sobolev space $W^{p,r}(\mathbb{R}^d)$ and for all $\zeta > (p + \tilde{p} + d/r^*)/m$, there exists a universal constant $C \geqslant 1$ such that

$$\|g - g_n\|_{W^{p,r}(\mathbb{R}^d)} \leqslant C\mathfrak{C}_{h,m\zeta,p+\tilde{p}+d/r^*}(K_{g,p,m} n^{-2h/\zeta} + K_\mu n^{-h+h(p+\tilde{p}+d/r^*)/(\zeta m)}). \tag{1.30}$$

with $\mathfrak{C}_{h,\xi,u} = 2^{h+u}(1 - 2^{-\xi+u})^{-1}$.

Proof. For $k, n \in \mathbb{N}$, we introduce

$$n_k = \min\{n; n^h \geqslant 2^{\zeta km}\}, \quad \text{and} \quad k_n = \min\{k \in \mathbb{N}; n_k \geqslant n\}$$

First, we notice that $n_{k_n - 1} < n \leqslant n_{k_n}$. Moreover, if we define $C_2 = 2^{\zeta m}, C_1 = 2^{-h}$, we have

$$C_1 n^h \leqslant 2^{\zeta k_n m} \leqslant C_2 n^h. \tag{1.31}$$

Indeed $n^h \geqslant n^h_{k_n - 1}$ which gives C_2. In order to obtain C_1, we notice that $n_{k_n} \leqslant 1 + 2^{\zeta k_n m/h}$. Now, we fix $n \in \mathbb{N}$ and for $k \in \mathbb{N}^*$, we define

$$\tilde{g}_k = 0 \text{ if } k < k_n \text{ and } \tilde{g}_k = g_{n_k} - g_n, \text{ if } k \geqslant k_n$$

and $\nu(dx) = \mu(dx) - g_n(x)dx, \nu_k(dx) = \tilde{g}_k(x)dx$. Using Proposition 2.5 and Theorem 2.6 in [8], it follows that

$$\|g - g_n\|_{W^{p,r}(\mathbb{R}^d)} \leqslant \sum_{k=1}^{\infty} 2^{k(p+\tilde{p}+d/r^*)} d_{\tilde{p}}(\nu, \nu_k) + \sum_{k=1}^{\infty} 2^{-2mk} \|\tilde{g}_k\|_{p+2m,2m,\tilde{p}} =: T_1 + T_2$$

First, we estimate T_1. If $k < k_n$, we have $\nu_k = 0$ so that $d_{\tilde{p}}(\nu, \nu_k) = d_{\tilde{p}}(\nu, 0) = d_{\tilde{p}}(\mu, \mu_{g_n}) \leqslant K_\mu / n^h$. On the other and, if $k \geqslant k_n$, we have $d_{\tilde{p}}(\nu, \nu_k) = d_{\tilde{p}}(\mu, \mu_{g_{n_k}}) \leqslant K_\mu n_k^{-h} \leqslant K_\mu 2^{-km\zeta}$. Using (1.31) together with all $\zeta > (p + \tilde{p} + d/r^*)/m$, it follows that

$$\begin{aligned} T_1 \leqslant & K_\mu 2^{k_n(p+\tilde{p}+d/r^*)} n^{-h} + (1 - 2^{-m\zeta+p+\tilde{p}+d/r^*})^{-1} K_\mu 2^{-k_n(m\zeta-p-\tilde{p}-d/r^*)} \\ \leqslant & 2(1 - 2^{-m\zeta+p+\tilde{p}+d/r^*})^{-1} C_2^{(p+\tilde{p}+d/r^*)/(\zeta m)} C_1^{-1} K_\mu / n^{h(1-(p+\tilde{p}+d/r^*)/(\zeta m))}. \end{aligned}$$

Now, we estimate T_2. Using (1.29) and (1.31) again, we have

$$T_2 \leqslant 2K_{g,p,m} \sum_{k=k_n}^{\infty} 2^{-2mk} \leqslant 2(1 - 2^{-2m})^{-1} C_1^{-1} K_{g,p,m} n^{-2h/\zeta},$$

and since $m \geqslant 1$, the proof is completed. □

We come back to our framework. We fix $R > 0$, $t \in \pi^{2S,T}_{T,n}$. We consider a function $\Phi_R \in \mathcal{C}^\infty_b(\mathbb{R}^d \times \mathbb{R}^d)$ such that $\mathbb{1}_{\overline{B}_R(x_0,y_0)}(x, y) \leqslant \Phi_R(x, y) \leqslant \mathbb{1}_{B_{R+1}(x_0,y_0)}$ and we denote

$$g^{n,R}_t(x, y) = \Phi_R(x, y)\overline{p}^n_t(x, y).$$

We use the result above for the sequence $g_n := g^{n,R}_t$, $n \in \mathbb{N}$ and $\mu(dx, dy) = \Phi_R(x, y)P_t(x, dy)dx$. In our specific case (1.24) and (1.27) give $d_0(\mu, \mu_{g_n}) \leqslant CS^{-\eta(q)} n^{-h}$. Since we have also (1.26), it follows that (1.29) hold with $K_\mu = CS^{-\eta(q)}$ and $K_{g,p,m} = CS^{-\eta(p+2m+2d)}$. We deduce from Proposition 1.2.4 that $\Phi_R(x,y)P_t(x,dy)dx = \mu(dx,dy) = g(x,y)dxdy$ with $g \in W^{p,r}(\mathbb{R}^d)$. Moreover, using Sobolev's embedding theorem, for $\zeta > (p+d/r^*)/m$ and $u \leqslant p - d/r$ we have

$$\|g - g_n\|_{u,\infty} \leqslant C\|g - g_n\|_{W^{p,r}(\mathbb{R}^d)} \leqslant C\mathfrak{C}_{h,m\zeta,p+d/r^*}(S^{-\eta(p+2m+2d)}n^{-2h/\zeta} + S^{-\eta(q)}n^{-h+h(p+d/r^*)/(\zeta m)}).$$

We take $u = |\alpha| + |\beta|$, $r = d$, $p = u+1$ and $m = \lceil (1-\varepsilon)(u+d)/(2\varepsilon) \rceil$ and put $\zeta = 2/(1-\epsilon)$. In this case $\zeta \geqslant (p+d/r^*)/m + 2$ and we obtain

$$\|g - g_n\|_{|\alpha|+|\beta|,\infty} \leqslant C2^{h+u+d}(S^{-\eta(u+2d+1+2\lceil (1-\varepsilon)(u+d)/(2\varepsilon)\rceil)}n^{-h(1-\varepsilon)} + S^{-\eta(q)}n^{-h(1-\varepsilon)}).$$

□

1.3 Integration by parts using a splitting method

In this section, we aim to build some integration by part formulas in order to prove the regularization properties. This kind of formulas is widely studied in Malliavin calculus for the Gaussian framework. However, since we are interested in random variables with form (1.1), where the random variables laws of $Z_k, k \in \mathbb{N}^*$ are arbitrary (and thus not Gaussian) the standard Malliavin calculus is not adapted anymore. Therefore, we whether develop a finite dimensional differential calculus which happens to be well suited to our framework as soon as Z_k involves a regular part.

Notice that in Section 1.6, we present the standard Malliavin calculus from the simple functionals perspective which is the approach that has inspired the finite differential calculus developed in this paper.

Concretely, we consider a sequence of independent random variables $Z_k = (Z_k^1, \dots, Z_k^N) \in \mathbb{R}^N$, $k \in \{1,\dots,n\}$ and we denote $Z = (Z_1,\dots,Z_n)$. The number n is fixed throughout this section (so there is no asymptotic procedure going on even if n is large in concrete applications since we are interested in estimating the error as $n \to \infty$). We aim to build integration by parts formulas based on the random vectors Z. The basic required assumption to obtain those formulas is the following: There exists $z_* = (z_{*,k})_{k\in\mathbb{N}^*}$ taking its values in $\mathbb{R}^N$ and $\varepsilon_*, r_* > 0$ such that for every Borel set $A \subset \mathbb{R}^N$ and every $k \in \{1,\dots,n\}$

$$L_{z_*}(\varepsilon_*, r_*) \qquad \mathbb{P}(Z_k \in A) \geqslant \varepsilon_* \lambda(A \cap B_{r_*}(z_{*,k})) \tag{1.32}$$

where λ is the Lebesgue measure on $\mathbb{R}^N$. We also define

$$M_p(Z) := 1 \vee \sup_{k\in\{1,\dots,n\}} \mathbb{E}[|Z_k|^p] \tag{1.33}$$

and assume that $M_p(Z) < \infty$ for every $p \geqslant 1$.
It is easy to check that (1.32) holds if and only if there exists some non negative measures μ_k with total mass $\mu_k(\mathbb{R}^N) < 1$ and a lower semi-continuous function $\varphi \geqslant 0$ such that $\mathbb{P}(Z_k \in dz) = \mu_k(dz) + \varphi(z - z_{*,k})dz$. Notice that the random variables $Z_1,\dots,Z_n$ are not assumed to be identically distributed. However, the fact that $r_* > 0$ and $\varepsilon_* > 0$ are the same for all k represents a mild substitute of this property. In order to construct φ we have to introduce the following function: For $v > 0$, set $\varphi_v : \mathbb{R}^N \to \mathbb{R}$ defined by

$$\varphi_v(z) = \mathbb{1}_{|z|\leqslant v} + \exp\Big(1 - \frac{v^2}{v^2 - (|z|-v)^2}\Big)\mathbb{1}_{v<|z|<2v}. \tag{1.34}$$

Then $\varphi_v \in \mathcal{C}_b^\infty(\mathbb{R}^N)$, $0 \leqslant \varphi_v \leqslant 1$ and we have the following crucial property: For every $p, k \in \mathbb{N}$ there exists a universal constant $C_{q,p}$ such that for every $z \in \mathbb{R}^N$, $q \in \mathbb{N}$ and $i_1, \ldots, i_q \in \{1, \ldots, N\}$, we have

$$\varphi_v(z)|\frac{\partial^q}{\partial z^{i_1} \cdot \partial z^{i_q}}(\ln \varphi_v)(z)|^p \leqslant \frac{C_{q,p}}{v^{pq}}, \tag{1.35}$$

with the convention $\ln \varphi_v(z) = 0$ for $|z| \geqslant 2v$. As an immediate consequence of (1.32), for every non negative function $f : \mathbb{R}^N \to \mathbb{R}_+$

$$\mathbb{E}[f(Z_k)] \geqslant \varepsilon_* \int_{\mathbb{R}^N} \varphi_{r_*/2}(\ z - z_{*,k}\ )f(z)dz.$$

By a change of variable

$$\mathbb{E}[f(\frac{1}{\sqrt{n}}Z_k)] \geqslant \varepsilon_* \int_{\mathbb{R}^N} n^{N/2}\varphi_{r_*/2}(\sqrt{n}(z - \frac{z_{*,k}}{\sqrt{n}}))f(z)dz. \tag{1.36}$$

We denote

$$m_* = \varepsilon_* \int_{\mathbb{R}^N} \varphi_{r_*/2}(z)dz = \varepsilon_* \int_{\mathbb{R}^N} \varphi_{r_*/2}(z - z_{*,k})dz$$

and

$$\phi_n(z) = n^{N/2}\varphi_{r_*/2}(\sqrt{n}z)$$

and we notice that $\int \phi_n(z)dz = m_*\varepsilon_*^{-1}$.
We consider a sequence of independent random variables $\chi_k \in \{0,1\}$, $U_k, V_k \in \mathbb{R}^N$, $k \in \{1, \ldots, n\}$ with laws given by

$$\begin{aligned}
\mathbb{P}(\chi_k &= 1) = m_*, \qquad \mathbb{P}(\chi_k = 0) = 1 - m_*, \\
\mathbb{P}(U_k &\in dz) = \frac{\varepsilon_*}{m_*}\phi_n(z - \frac{z_{*,k}}{\sqrt{n}})dz, \\
\mathbb{P}(V_k &\in dz) = \frac{1}{1 - m_*}(\mathbb{P}(\frac{1}{\sqrt{n}}Z_k \in dz) - \varepsilon_*\phi_n(z - \frac{z_{*,k}}{\sqrt{n}})dz).
\end{aligned} \tag{1.37}$$

Notice that (1.36) guarantees that $\mathbb{P}(V_k \in dz) \geqslant 0$. Then a direct computation shows that

$$\mathbb{P}(\chi_k U_k + (1 - \chi_k)V_k \in dz) = \mathbb{P}(\frac{1}{\sqrt{n}}Z_k \in dz). \tag{1.38}$$

This is the splitting procedure for $\frac{1}{\sqrt{n}}Z_k$. Now on we will work with this representation of the law of $\frac{1}{\sqrt{n}}Z_k$. So, we always take

$$\frac{1}{\sqrt{n}}Z_k = \chi_k U_k + (1 - \chi_k)V_k.$$

Remark 1.3.1. *The above splitting procedure has already been widely used in the litterature: In [60] and [51], it is used in order to prove convergence to equilibrium of Markov processes. In [16], [17] and [69], it is used to study the Central Limit Theorem. Last but not least, in [59], the above splitting method (with $\mathbb{1}_{B_{r_*}(z_{*,k})}$ instead of $\phi_n(z - \frac{z_{*,k}}{\sqrt{n}})$) is used in a framework which is similar to the one in this paper.*

In the following, we will denote $\chi=(\chi_1,\ldots,\chi_n),\ U=(U_1,\ldots,U_n)$ and $V=(V_1,\ldots,V_n)$ and we will consider the class of random variables:

$$\mathcal{S}=\{F=f(\chi,U,V): f \text{ is measurable and } u\to f(\chi,u,v)\in\mathcal{C}_b^\infty(\mathbb{R}^n\times\mathbb{R}^N),\forall\chi,v\}. \tag{1.39}$$

For a multi index $\alpha=(\alpha_1,\ldots,\alpha_q)$ with $\alpha_j=(k_j,i_j),\ k_j\in\{1,\ldots,n\},\ i_j\in\{1,\ldots,N\}$, we denote $|\alpha|=q$ the length of α and

$$\partial_u^\alpha f(\chi,u,v)=\frac{\partial^q}{\partial u_{k_1}^{i_1}\cdots\partial u_{k_q}^{i_q}}f(\chi,u,v).$$

We construct now a differential calculus based on the laws of the random variables $U_k,\ k=1,\ldots,n$ which mimics the Malliavin calculus, following the ideas from [9], [6] and [7]. In order to be self contained, we shortly present the results that we need. For $F=f(\chi,U,V)\in\mathcal{S}$ we define the Malliavin derivatives

$$D_{(k,i)}F=\chi_k\frac{1}{\sqrt{n}}\frac{\partial F}{\partial U_k^i}=\chi_k\frac{1}{\sqrt{n}}\frac{\partial f}{\partial u_k^i}(\chi,U,V),\qquad k=1,\ldots,n,\ \ i=1,\ldots,N. \tag{1.40}$$

We denote by $\langle\cdot,\cdot\rangle$ the usual scalar product on $\mathbb{R}^N\times\mathbb{R}^n$. The Malliavin covariance matrix for a multi dimensional functional $F=(F^1,\ldots,F^d)$ is defined as

$$\sigma_F^{i,j}=\left\langle DF^i,DF^j\right\rangle=\sum_{k=1}^n\sum_{r=1}^N D_{(k,r)}F^i\times D_{(k,r)}F^j,\qquad i,j=1,\ldots,d. \tag{1.41}$$

The higher order derivatives are defined by iterating D:

$$D_\alpha F=D_{\alpha_1}\cdots D_{\alpha_m}F.$$

Now we define the Ornstein Uhlenbeck operator $L:\mathcal{S}\to\mathcal{S}$. We denote

$$\Gamma_k=\ln\phi_n(U_k-\frac{z_{*,k}}{\sqrt{n}})\in\mathcal{S}$$

and we notice that

$$\begin{aligned}D_{(k,i)}\Gamma_k &= \frac{1}{\sqrt{n}}\chi_k\partial_{u_k^i}\ln\phi_n(U_k-\frac{z_{*,k}}{\sqrt{n}})=\frac{1}{\sqrt{n}}\chi_k\partial_{u_k^i}\ln\phi_n(u_k-\frac{z_{*,k}}{\sqrt{n}})|_{u_k=U_k}\\ &= \chi_k\partial_{z^i}\ln\varphi_{r_*/2}(z)\big|_{z=\sqrt{n}(U_k-\frac{z_{*,k}}{\sqrt{n}})}.\end{aligned}$$

Finally, we define

$$-LF=\sum_{k=1}^n\sum_{i=1}^N D_{(k,i)}D_{(k,i)}F+\sum_{k=1}^n\sum_{i=1}^N D_{(k,i)}F\times D_{(k,i)}\Gamma_k.$$

Remark 1.3.2. *The basic random variables in our calculus are $Z_k,\ k=1,\ldots,n$ so we precise the way in which the differential operators act on them. Since $Z_k=\sqrt{n}\chi_kU_k+\sqrt{n}(1-\chi_k)V_k$, it follows that*

$$D_{(m,j)}Z_k^i = \chi_k\delta_{m,k}\delta_{i,j}, \tag{1.42}$$

$$LZ_k^i = -\chi_k\partial_{z^i}\ln\varphi_{r_*/2}(z)\big|_{z=\sqrt{n}(U_k-\frac{z_{*,k}}{\sqrt{n}})}. \tag{1.43}$$

where $\delta_{i,j}=1$ if $i=j$ and 0 if $i\neq j$, stands for the Kroenecker symbol.

In our framework, the duality formula in Malliavin calculus reads as follows: For each $F, G \in \mathcal{S}$

$$\mathbb{E}[FLG] = \mathbb{E}[\langle DF, DG \rangle] = \mathbb{E}[GLF]. \tag{1.44}$$

This follows immediately using the independence structure and standard integration by parts on $\mathbb{R}^N$: Indeed, if $f, g \in \mathcal{C}_b^1(\mathbb{R}^N)$ and $k \in \{1, \ldots, n\}$, then

$$\begin{aligned}
&\sum_{i=1}^N \mathbb{E}[\partial_{u_k^i} f(U_k) \partial_{u_k^i} g(U_k)] \\
= \;& \frac{\varepsilon_*}{m_*} \sum_{i=1}^N \int_{\mathbb{R}^N} \partial_{u_k^i} f(u) \partial_{u_k^i} g(u) \phi_n(u - \frac{z_{*,k}}{\sqrt{n}}) du \\
= \;& -\frac{\varepsilon_*}{m_*} \sum_{i=1}^N \int_{\mathbb{R}^N} f(u) (\partial^2_{u_k^i} g(u) + \partial_{u_k^i} g(u) \frac{\partial_{u_k^i} \phi_n(u - \frac{z_{*,k}}{\sqrt{n}})}{\phi_n(u - \frac{z_{*,k}}{\sqrt{n}})}) \phi_n(u - \frac{z_{*,k}}{\sqrt{n}}) du \\
= \;& -\mathbb{E}\Big[f(U_k) \sum_{i=1}^N \partial^2_{u_k^i} g(U_k) + \partial_{u_k^i} g(U_k) \partial_{u_k^i} \ln \phi_n(U_k - \frac{z_{*,k}}{\sqrt{n}})\Big].
\end{aligned}$$

It follows that

$$\begin{aligned}
&\sum_{k=1}^n \sum_{i=1}^N \mathbb{E}[D_{(k,i)} F \times D_{(k,i)} G] \\
= \;& \frac{1}{n} \sum_{k=1}^n \sum_{i=1}^N \mathbb{E}[\chi_k \partial_{u_k^i} f(\chi, U, V) \times \partial_{u_k^i} g(\chi, U, V)] \\
= \;& -\mathbb{E}\Big[f(\chi, U, V) \sum_{k=1}^n \chi_k \sum_{i=1}^N \frac{1}{n} \partial^2_{u_k^i} g(\chi, U, V) + \frac{1}{\sqrt{n}} \partial_{u_k^i} g(\chi, U, V) \frac{1}{\sqrt{n}} \partial_{u_k^i} \ln \phi_n(U_k - \frac{z_{*,k}}{\sqrt{n}})\Big] \\
= \;& -\mathbb{E}\Big[f(\chi, U, V) \sum_{k=1}^n \sum_{i=1}^N D_{(k,i)} D_{(k,i)} G + D_{(k,i)} G D_{(k,i)} \Gamma_k\Big] \\
= \;& \mathbb{E}[FLG],
\end{aligned}$$

which is exactly (1.44). We have the following standard chain rule: For $\phi \in \mathcal{C}^1(\mathbb{R}^d)$ and $F \in \mathcal{S}^d$

$$D\phi(F) = \sum_{j=1}^d \partial_j \phi(F) DF^j. \tag{1.45}$$

Moreover, one may prove, using (1.45) and the duality relation (or direct computation), that

$$L\phi(F) = \sum_{j=1}^d \partial_j \phi(F) LF^j + \sum_{i,j=1}^d \partial_i \partial_j \phi(F) \left\langle DF^i, DF^j \right\rangle. \tag{1.46}$$

In particular for $F, G \in \mathcal{S}$,

$$L(FG) = FLG + GLF + 2 \langle DF, DG \rangle. \tag{1.47}$$

We are now able to give the Malliavin integration by parts formula:

Theorem 1.3.1. *Let $F \in \mathcal{S}^d$ and $G \in \mathcal{S}$ be such that $\mathbb{E}[(\det \sigma_F)^{-p}] < \infty$ for every $p \geqslant 1$. We denote $\gamma_F = \sigma_F^{-1}$. Then for every $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ and every $i = 1, \ldots, d$*

$$\mathbb{E}[\partial_i \phi(F) G] = \mathbb{E}[\phi(F) H_i(F, G)] \tag{1.48}$$

with

$$-H(F,G) = G\gamma_F LF + \langle D(G\gamma_F), DF\rangle \tag{1.49}$$

and

$$H_i(F,G) = -\sum_{j=1}^{d} G\gamma_F^{i,j} LF^j + \langle D(G\gamma_F^{i,j}), DF^j\rangle.$$

Moreover, for every multi index $\alpha = (\alpha_1, \ldots, \alpha_m) \in \{1, \ldots, d\}^m$

$$\mathbb{E}[\partial_\alpha \phi(F) G] = \mathbb{E}[\phi(F) H_\alpha(F,G)] \tag{1.50}$$

with $H_\alpha(F,G)$ *defined by the recurrence relation* $H_{(\alpha_1,\cdots,\alpha_m)}(F,G) = H_{\alpha_m}(F, H_{(\alpha_1,\ldots,\alpha_{m-1})}(F,G))$.

Proof. Using the chain rule $D\phi(F) = \nabla\phi(F) DF$ we have

$$\langle D\phi(F), DF\rangle = \nabla\phi(F)\langle DF, DF\rangle = \nabla\phi(F)\sigma_F.$$

It follows that $\nabla\phi(F) = \gamma_F \langle D\phi(F), DF\rangle$. Then, using (1.47) and the duality formula (1.44),

$$\begin{aligned}\mathbb{E}[G\nabla\phi(F)] &= \mathbb{E}[G\gamma_F \langle D\phi(F), DF\rangle] = \frac{1}{2}\mathbb{E}[G\gamma_F(L(\phi(F)F) - \phi(F)LF - FL\phi(F))] \\ &= \frac{1}{2}\mathbb{E}[\phi(F)(FL(G\gamma_F) - G\gamma_F LF - L(G\gamma_F F))].\end{aligned}$$

We use once again (1.47) in order to obtain $H(F,G)$ in (1.141). □

We give now estimates of the weights $H_\alpha(F,G)$ which appear in the above integration by parts formulas. We will work with the norms:

$$|F|_{1,q}^2 = \sum_{1\leqslant|\alpha|\leqslant q} |D_\alpha F|^2, \qquad |F|_q^2 = |F|^2 + |F|_{1,q}^2, \tag{1.51}$$

and

$$\begin{aligned}\|F\|_{1,q,p} &= \big\||F|_{1,q}\big\|_p = \mathbb{E}[|F|_{1,q}^p]^{1/p} \\ \|F\|_{q,p} &= \|F\|_p + \big\||F|_{1,q}\big\|_p.\end{aligned} \tag{1.52}$$

Proposition 1.3.1. *For each* $m, q \in \mathbb{N}$, *there exists a universal constant* $C \geqslant 1$ *(depending on* d, m, q *only) such that for every multi index* α *with* $|\alpha| \leqslant q$ *and every* $F \in \mathcal{S}^d$ *and* $G \in \mathcal{S}$ *on has*

$$|H_\alpha(F,G)|_m \leqslant C(1 \vee (\det\sigma_F)^{-1})^{q(q+m+1)}(1 + |F|_{1,m+q+1}^{2dq(q+m+2)} + |LF|_{m+q-1}^{2q})|G|_{m+q}. \tag{1.53}$$

The proof is long but straightforward so we skip it. The reader may find the detailed proof in [9] and in [6], Theorem 3.4.
We end this section with an estimate of $\|LZ_k^i\|_{q,p}$:

Lemma 1.3.1. *We have the following properties.*

A. *For every* $k = 1, \ldots, n$ *and* $i = 1, \ldots, N$, *we have*

$$\mathbb{E}[LZ_k^i] = 0. \tag{1.54}$$

B. *For every* $q \in \mathbb{N}$ *and* $p \geqslant 2$ *there exists a constant* C *depending on* q, p *only*

$$\|LZ_k^i\|_{q,p} \leqslant \frac{Cm_*^{1/p}}{r_*}(1 + r_*^{-q}) \tag{1.55}$$

Proof. **A**. Using the duality relation we have $\mathbb{E}[1 \times LZ_k^i] = \mathbb{E}[\langle D1, DZ_k^i \rangle] = 0$. In order to prove **B** we recall (see (1.43)) that

$$LZ_k^i = -\chi_k \partial_i (\ln \varphi_{r_*/2})(\sqrt{n}(U_k - \frac{z_{*,k}}{\sqrt{n}})).$$

Let $\Lambda_{k,q}$ be the set of the multi-index $\alpha = (\alpha_1, \dots, \alpha_q)$ such that $\alpha_j = (k, i_j)$. Notice that for a multi-index α of length q, such that $\alpha \notin \Lambda_{k,q}$, we have $D_\alpha LZ_k^i = 0$. Suppose now that $\alpha \in \Lambda_{k,q}$ and let $\overline{\alpha} = (i_1, \dots, i_q, i)$. It follows

$$D_\alpha LZ_k^i = -\chi_k \partial_{\overline{\alpha}} (\ln \varphi_{r_*/2})(\sqrt{n}(U_k - \frac{z_{*,k}}{\sqrt{n}})).$$

Since the function $\varphi_{r_*/2}$ is constant on $B_{r_*/2}(0)$ and on $\mathbb{R}^d \setminus \overline{B}_{r_*}(0)$, using (1.35), we obtain

$$\begin{aligned}
\|D_\alpha LZ_k^i\|_p^p &= \frac{\varepsilon_* \|\chi_k\|_p^p}{m_*} \int_{\mathbb{R}^N} n^{N/2} |\partial_{\overline{\alpha}} (\ln \varphi_{r_*/2})(\sqrt{n}(u - \frac{z_{*,k}}{\sqrt{n}}))|^p \varphi_{r_*/2}(\sqrt{n}(u - \frac{z_{*,k}}{\sqrt{n}})) du \\
&= \frac{\varepsilon_* \|\chi_k\|_p^p}{m_*} \int_{r_*/2 \leqslant |u| \leqslant r_*} |\partial_{\overline{\alpha}} (\ln \varphi_{r_*/2})(u)|^p \varphi_{r_*/2}(u) du \\
&\leqslant \frac{C_{q+1,p} m_*}{r_*^{p(q+1)}}.
\end{aligned}$$

and then

$$\|LZ_k^i\|_{q,p} \leqslant C \sup_{l \leqslant q} \sup_{\alpha \in \Lambda_{k,l}} \|D_\alpha LZ_k^i\|_p \leqslant \frac{C m_*^{1/p}}{r_*} (1 + r_*^{-q}).$$

□

1.3.1 Localizaton

We have seen in Proposition 1.3.1 that we can bound the Sobolev norms of the weight which appear in the integration by part formula (1.142). In order to obtain the regularization properties, we will have to bound the moments of those Sobolev norms or more particularly, the moments of the terms which appear in the right hand side of (1.53). However, in many cases it is cumbersome to estimate $\mathbb{E}[(\det \sigma_F)^{-p}]$, $p \in \mathbb{N}$. The method adopted in this paper comes down to localize the calculus when $\det \sigma_F$ does not belong to a neighborhood of zero. Then, we will prove a similar property as (1.24) and we will obtain the convergence in total variation distance. More specifically, when $F = X^n$, we will have to localize the random variables Z_k and χ_k which appear in (1.1) with the decomposition (1.38). We introduce a suited framework to treat this problem.

In the following, we will not work under $\mathbb{P}$, but under a localized probability measure which we define now. We fix $S > 0$ such that $S \leqslant T$ and we consider the set

$$\Lambda_S = \{ \frac{1}{\lfloor Sn/T \rfloor} \sum_{k=1}^{\lfloor Sn/T \rfloor} \chi_k \geqslant \frac{m_*}{2} \}. \tag{1.56}$$

Using Hoeffding's inequality and the fact that $\mathbb{E}[\chi_k] = m_*$, it can be checked that

$$\mathbb{P}(\Lambda_S^c) \leqslant \exp(-m_*^2 \lfloor Sn/T \rfloor / 2). \tag{1.57}$$

We consider also the localization function $\varphi_{n^{1/4}/2}$, defined in (1.34), and we construct the random variable

$$\Theta = \Theta_{S,n} = \mathbb{1}_{\Lambda_S} \times \prod_{k=1}^{n} \varphi_{n^{1/4}/2}(Z_k). \tag{1.58}$$

Since Z_k has finite moments of any order, the following inequality can be shown: For every $l \in \mathbb{N}$ there exists C such that

$$\mathbb{P}(\Theta_{S,n} = 0) \leqslant \mathbb{P}(\Lambda_M^c) + \sum_{k=1}^{n} \mathbb{P}(|Z_k| \geqslant n^{1/4}) \leqslant \exp(-m_*^2 \lfloor Sn/T \rfloor /2) + \frac{M_{4(l+1)}(Z)}{n^l}. \tag{1.59}$$

We define the probability measure

$$d\mathbb{P}_\Theta = \frac{1}{\mathbb{E}[\Theta]} \Theta d\mathbb{P}. \tag{1.60}$$

Corollary 1.3.1. *Let $F \in \mathcal{S}^d$ and $G \in \mathcal{S}$ be such that $\mathbb{E}_\Theta[(\det \sigma_F)^{-p}] < \infty$ for every $p \geqslant 1$. We denote $\gamma_F = \sigma_F^{-1}$. Then, for every $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ and every $i = 1, \ldots, d$*

$$\mathbb{E}_\Theta[\partial_i \phi(F) G] = \mathbb{E}_\Theta[\phi(F) H_i^\Theta(F, G)] \tag{1.61}$$

with

$$-H^\Theta(F, G) = G\gamma_F LF + \langle D(G\gamma_F, DF\rangle + G\gamma_F \langle D \ln \Theta, DF\rangle$$

and

$$H_i^\Theta(F, G) = -\sum_{j=1}^{d} G\gamma_F^{i,j} LF^j + \langle D(G\gamma_F^{i,j}), DF^j\rangle + G\gamma_F^{i,j} \left\langle D \ln \Theta, DF^j \right\rangle.$$

And for every multi index $\alpha = (\alpha_1, \ldots, \alpha_m) \in \{1, \ldots, d\}^m$,

$$\mathbb{E}_\Theta[\partial_\alpha \phi(F) G] = \mathbb{E}_\Theta[\phi(F) H_\alpha^\Theta(F, G)], \tag{1.62}$$

with $H_\alpha^\Theta(F, G)$ defined by the recurrence relation $H^\Theta_{(\alpha_1, \cdots, \alpha_m)}(F, G) = H^\Theta_{\alpha_m}(F, H^\Theta_{(\alpha_1, \cdots, \alpha_{m-1})}(F, G))$, and the convention $\ln(\Theta) = 0$ for $\Theta = 0$. Moreover there exists an universal constant C such that for every multi index α with $|\alpha| = q$

$$\mathbb{E}_\Theta[|H_\alpha^\Theta(F, G)|_m^p] \leqslant CC_{q,\Theta}(F, G), \tag{1.63}$$

with

$$\begin{aligned} C_{q,\Theta}(F, G) &= \mathbb{E}_\Theta[(1 \vee (\det \sigma_F)^{-1})^{2pq(q+m+1)}]^{1/2} \\ &\quad \times (1 + \mathbb{E}_\Theta[|F|_{1,m+q+1}^{8pqd(q+m+2)}]^{1/4} + \mathbb{E}_\Theta[|LF|_{m+q-1}^{8pq}]^{1/4}) \mathbb{E}_\Theta[|G|_{m+q}^{4p}]^{1/4}. \end{aligned} \tag{1.64}$$

Proof. Using (1.141) with G replaced by $G\Theta$ we obtain $\mathbb{E}[\partial_i \phi(F) G\Theta] = \mathbb{E}[\phi(F) \overline{H}_i]$ with

$$\overline{H} = -\Theta G\gamma_F LF - \langle D(\Theta G\gamma_F), DF\rangle = \Theta H(F, G) - G\gamma_F \langle D\Theta, DF\rangle.$$

It follows that

$$\begin{aligned} \mathbb{E}_\Theta[\partial_i \phi(F) G] &= \frac{1}{\mathbb{E}[\Theta]} \mathbb{E}[\partial_i \phi(F) G\Theta] = \frac{1}{\mathbb{E}[\Theta]} \mathbb{E}[\phi(F)(\Theta H_i(F, G) - G \sum_{j=1}^{d} \gamma_F^{i,j} \left\langle D\Theta, DF^j \right\rangle] \\ &= \mathbb{E}_\Theta[\phi(F)(H_i(F, G) - G \sum_{j=1}^{d} \gamma_F^{i,j} \left\langle D \ln \Theta, DF^j \right\rangle]. \end{aligned}$$

So (1.61) is proved and (1.62) follows by recurrence. Moreover

$$\mathbb{E}_\Theta[|G\sum_{j=1}^{d}\gamma_F^{i,j}\left\langle D\ln\Theta, DF^j\right\rangle|^p] \leqslant C\mathbb{E}_\Theta[|D\ln(\Theta)^{4p}]^{1/4}\mathbb{E}_\Theta[|\gamma_F|^{4p}]^{1/4}\mathbb{E}_\Theta[|DF|^{4p}]^{1/4}\mathbb{E}_\Theta[|G|^{4p}]^{1/4}.$$

Notice that by (1.35) we have

$$\mathbb{E}_\Theta[|D\ln\Theta|^{4p}]^{1/4p} \leqslant C/n^{1/4}.$$

Then (1.63) follows from (1.53). □

1.4 Convergence results for a class of Markov Chain

Now we have introduced the integration by parts formulas which are adapted to our study, we are in a position to prove the regularization properties. In order to do it, we have to bound the miscellaneous terms which appear in the right hand side of (1.64). This section is devoted to the estimation of those terms. We will treat separately the estimation of the norm of the inverse of the covariance matrix from the other terms. Indeed, this study requires localization techniques which are not necessary in order to bound the Sobolev norms of the others terms. Then we will give the regularization properties and the total variation convergence results that follow from those estimates.

Throughout this section, $n\in\mathbb{N}^*$ will still be fixed and will be the number of time step between 0 and T and also the number of increments that we consider in our abstract Malliavin calculus. We consider two sequences of independent random variables $Z_{k+1}\in\mathbb{R}^N, \kappa_k\in\mathbb{R},\ k\in\mathbb{N}$ and we assume that $Z_k,\ k\in\mathbb{N}^*$, are centered and verify (1.32) and (1.33).
We suppose that, there exists $C\geqslant 1$ such that $\sup_{k\in\mathbb{N}^*}\delta_k^n\leqslant C/n$ and we construct the $\mathbb{R}^d$ valued Markov chain $(X_t^n)_{t\in\pi_{T,n}}$ in the following way:

$$X^n_{t^n_{k+1}} = \psi(\kappa_k, X^n_{t^n_k}, \frac{Z_{k+1}}{\sqrt{n}}, \delta^n_{k+1}), \qquad k\in\mathbb{N} \tag{1.65}$$

where

$$\psi\in\mathcal{C}^\infty(\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^N\times\mathbb{R}_+;\mathbb{R}^d) \quad\text{and}\quad \psi(\kappa,x,0,0)=x. \tag{1.66}$$

We introduce the norm

$$\|\psi\|_{1,r,\infty} = 1\vee\sum_{|\alpha|=0}^{r}\sum_{|\beta|+|\gamma|=1}^{r-|\alpha|}\|\partial_x^\alpha\partial_z^\beta\partial_t^\gamma\psi\|_\infty. \tag{1.67}$$

Remark 1.4.1. *Notice that the random variables κ_k can be useful in concrete applications. Indeed, in the Ninomiya Victoir scheme, at each time step k, one throws a coin $\kappa_k\in\{1,-1\}$ and uses different forms for the function ψ according to the fact that κ_k is equal to 1 or to -1.*

Since the function ψ only needs to be measurable with respect to κ and that all our estimates will be done in terms of $\|\psi\|_{1,r,\infty}$, then without loss of generality, we can simplify the notations and denote

$$\psi_k(x,z,t)=\psi(\kappa_k,x,z,t).$$

Then, we slightly modify the definition (1.67) and instead, in the sequel, we will consider the norm

$$\|\psi\|_{1,r,\infty} = \sup_{k\in\mathbb{N}} \|\psi_k\|_{1,r,\infty} = 1 \vee \sup_{k\in\mathbb{N}} \sum_{|\alpha|=0}^{r} \sum_{|\beta|+|\gamma|=1}^{r-|\alpha|} \|\partial_x^\alpha \partial_z^\beta \partial_t^\gamma \psi_k\|_\infty, \tag{1.68}$$

with $(\psi_k)_{k\in\mathbb{N}}$ a sequence of functions that belong to $\mathcal{C}^r(\mathbb{R}^d \times \mathbb{R}^N \times \mathbb{R}_+; \mathbb{R}^d)$. It is worth noticing that all our results remain true if we replace the supremum over $k \in \mathbb{N}$ by the supremum over $k \in \mathbb{N}$ with $t_k^n < T$. However, for the sake of clarity, we will work with (1.68). Finally for $r \in \mathbb{N}^*$, we denote

$$\mathfrak{K}_r(\psi) = (1 + \|\psi\|_{1,r,\infty}) \exp(\|\psi\|_{1,3,\infty}^2). \tag{1.69}$$

We aim to give sufficient conditions under which the above Markov chain has the regularization property (1.25). In order to do it, we consider the following new representation of X^n. Let us introduce some notations. We denote

$$H_k = \frac{Z_k}{\sqrt{n}} = \chi_k U_k + (1-\chi_k) V_k.$$

Using a Taylor expansion of order one, we write

$$\begin{aligned} X^n_{t^n_{k+1}} &= X^n_{t^n_k} + \sum_{i=1}^N \partial_{z_i}\psi_k(X^n_{t^n_k}, 0, 0) H^i_{k+1} + \delta^n_{k+1} \int_0^1 \partial_t \psi_k(X^n_{t^n_k}, H_{k+1}, \lambda\delta^n_{k+1}) d\lambda \\ &+ \frac{1}{2} \sum_{i,j=1}^N H^i_{k+1} H^j_{k+1} \int_0^1 (1-\lambda) \partial_{z_i}\partial_{z_j} \psi_k(X^n_{t^n_k}, \lambda H_{k+1}, 0) d\lambda. \end{aligned}$$

We denote

$$a^i_k = \partial_{z_i}\psi_k(X^n_{t^n_k}, 0, 0),\ b^{i,j}_k = \int_0^1 (1-\lambda)\partial_{z_i}\partial_{z_j}\psi_k(X^n_{t^n_k}, \lambda H_{k+1}, 0) d\lambda,\ \tilde{b}_k = \int_0^1 \partial_t \psi_k(X^n_{t^n_k}, H_{k+1}, \lambda\delta^n_{k+1}) d\lambda,$$

and then, we write

$$X^n_{t^n_m} = x + \sum_{i=1}^N \sum_{k=0}^{m-1} a^i_k H^i_{k+1} + \frac{1}{2} \sum_{i,j=1}^N \sum_{k=0}^{m-1} b^{i,j}_k H^i_{k+1} H^j_{k+1} + \sum_{k=0}^{m-1} \tilde{b}_k \delta^n_{k+1}. \tag{1.70}$$

Moreover we denote by $X^n(x)$ the Markov chain which starts from x (i.e. $X^n_0(x) = x$) and we denote by $\partial^\alpha X^n$ the derivative with respect to the starting point x. We will use the results from the previous section for X^n. In order to do it we have to estimate the Sobolev norms of X^n:

Theorem 1.4.1. *For every $q, q' \in \mathbb{N}$ with $q \geqslant q'$, and $p \geqslant 2$ there exists $l \in \mathbb{N}^*$, $C \geqslant 1$ which depend on $r_*, \varepsilon_*, m_*, q, p$ and the moments of Z, but not on n,such that*

$$\sup_{t\in\pi^T_{T,n}} \sup_{0\leqslant|\alpha|\leqslant q-q'} \|\partial_x^\alpha X^n_t(x)\|_{q',p} \leqslant C\mathfrak{K}_{q+2}(\psi)^l, \tag{1.71}$$

$$\sup_{t\in\pi^T_{T,n}} \|LX^n_t\|_{q,p} \leqslant C\mathfrak{K}_{q+4}(\psi)^l, \tag{1.72}$$

where $\mathfrak{K}_r(\psi)$ is defined in (1.69) and is given by

$$\mathfrak{K}_r(\psi) = (1 + \|\psi\|_{1,r,\infty}) \exp(\|\psi\|_{1,3,\infty}^2).$$

The proof is long and technical so we postpone it to Section 1.4.4.

1.4.1 The Malliavin covariance matrix

We turn now to the covariance matrix. We will work under the probability $\mathbb{P}_\Theta$ defined in (1.60). We recall that $T>0$ and $n\in\mathbb{N}$ are given and we have denoted $\Lambda_S=\{\frac{1}{\lfloor nS/T\rfloor}\sum_{k=1}^{\lfloor nS/T\rfloor}\chi_k\geqslant\frac{m_*}{2}\}$. The localization random variable $\Theta_{S,n}$ is defined in (1.58) and we have proved in (1.59) that, for every $l\in\mathbb{N}$,

$$\mathbb{P}(\Theta_{S,n}=0)\leqslant\exp(-m_*^2\lfloor nS/T\rfloor/2)+\frac{M_{4(l+1)}(Z)}{n^l}.$$

We also have

$$\{\Theta_{S,n}\neq 0\}\subset\{\frac{1}{n}\sum_{k=1}^{\lfloor nS/T\rfloor}\chi_k\geqslant\frac{\lfloor nS/T\rfloor m_*}{2n}\}\cap\{|Z_k|\leqslant n^{1/4},\ k=1,\ldots,n\}.$$

Using the computational rules for $k\in\{0,\ldots,m-1\}$ and $m\leqslant n$, we obtain

$$D_{(k+1,i)}X^n_{t^n_m}=I_{k,i}+\sum_{l=k+1}^{m-1}J_lD_{(k+1,i)}X^n_{t^n_l} \tag{1.73}$$

with

$$\begin{aligned}
I_{k,i} &= \frac{1}{\sqrt{n}}\chi_{k+1}(a^i_k+\sum_{j=1}^N H^j_{k+1}b^{i,j}_k+\sum_{j,q=1}^N H^j_{k+1}H^q_{k+1}c^{i,j,q}_k+\delta^n_{k+1}\tilde{c}^i_k), &(1.74)\\
c^{i,j,q}_k &= \frac{1}{\sqrt{n}}\chi_{k+1}\int_0^1\lambda(1-\lambda)\partial_{z_i}\partial_{z_j}\partial_{z_q}\psi_k(X^n_{t^n_k},\lambda H_{k+1},0)d\lambda\\
\tilde{c}^i_k &= \int_0^1\partial_{z_i}\partial_t\psi_k(X^n_{t^n_k},H_{k+1},\lambda\delta^n_{k+1})d\lambda &(1.75)
\end{aligned}$$

and the $d\times d$ dimensional matrices J_l, defined by

$$J^{p,r}_l=J^{p,r}_{l,0}+\sum_{j=1}^N J^{p,r}_l(j)+\sum_{j,q=1}^N J^{p,r}_l(j,q)$$

with

$$\begin{aligned}
J^{p,r}_{l,0} &= \delta^n_{l+1}\int_0^1\partial_{x_p}\partial_t(\psi_l(X^n_{t^n_l},H_{l+1},\lambda\delta^n_{l+1}))_r d\lambda,\\
J^{p,r}_l(j) &= H^j_{l+1}\partial_{x_p}\partial_{z_j}(\psi_l(X^n_{t^n_l},0,0))_r,\\
J^{p,r}_l(j,q) &= H^j_{l+1}H^q_{l+1}\int_0^1(1-\lambda)\partial_{x_p}\partial_{z_j}\partial_{z_q}(\psi_l(X^n_{t^n_l},\lambda H_{l+1},0))_r d\lambda.
\end{aligned}$$

We first aim to express $D_{(k+1,i)}X^n_t$ using the variance of constants method. We consider the tangent flow $Y^n_t=\nabla_x X^n_t(x)$, $t\in\pi_{T,n}$, which is the $d\times d$ dimensional matrix solution of

$$Y^n_{t^n_m}=I+\sum_{l=0}^{m-1}J_lY^n_{t^n_l},$$

where I is the identity matrix. The explicit solution of the above equation is given by $Y^n_{t^n_m}=\prod_{k=0}^{m-1}(I+J_k)$. If each of the matrices $I+J_k,\ k=1,\ldots,m$, is invertible then, $Y^n_{t^n_m}$ is also invertible. On the set $\{\Theta_{t^n_m,n}\neq 0\}$, we have $|H_k|=|n^{-1/2}Z_k|\leqslant n^{-1/4}$ so that $\|J_k\|_\infty:=\sup_{i,j\leqslant d}\|J^{i,j}_k\|_\infty\leqslant 3\|\psi\|_{1,3,\infty}n^{-1/4}$. It follows that, among others, if $\|\psi\|_{1,3,\infty}n^{-1/4}\leqslant 1/6$, then the lower eigenvalue of $I+J_k$ is larger then $1/2$, so we have the invertibility property. We denote by $(\widehat{Y}_t^n)_{t\in\pi_{T,n}}$ the inverse of $(Y_t^n)_{t\in\pi_{T,n}}$ and it is easy to check that $\widehat{Y}^n$ solves the equation:

$$\widehat{Y}^n_{t^n_m} = I - \sum_{l=0}^{m-1} \widehat{Y}^n_{t^n_l}(I+J_l)^{-1}J_l.$$

The following representation of the Malliavin derivative, known as the "variance of constants method", is given by

$$\forall t \in \pi_{T,n}, t \geqslant t^n_{k+1} \qquad D_{(k+1,i)}X^n_t = Y^n_t \widehat{Y}^n_{t^n_{k+1}} I_{k,i}, \tag{1.76}$$

and is zero if $t < t^n_{k+1}$. We will use the following estimates.

Lemma 1.4.1. *Let $p \geqslant 2$. There exists a constant $C \geqslant 1$, which depends on p and T, such that the following holds. Suppose that n and $t \in \pi^{0,T}_{T,n}$ are sufficiently large in order to have*

$$\frac{3\|\psi\|_{1,3,\infty}}{n^{1/4}} + \frac{M_8(Z)}{n} + \exp(-m_*^2 nt/(2T)) \leqslant \frac{1}{2}. \tag{1.77}$$

Then,

$$\mathbb{E}_{\Theta_{t,n}}[\sup_{s\in\pi^T_{T,n}} \|Y^n_s\|^p] \leqslant 2\exp\left(C(M_{2p}(Z)^{2/p} + M_4(Z)^2)(1+\|\psi\|^2_{1,3,\infty})\right), \tag{1.78}$$

and

$$\mathbb{E}_{\Theta_{t,n}}[\sup_{s\in\pi^T_{T,n}} \|\widehat{Y}^n_s\|^p] \leqslant 2\exp\left(C(M_{2p}(Z)^{2/p} + M_4(Z)^2)(1+\|\psi\|^2_{1,3,\infty})\right), \tag{1.79}$$

with

$$\|Y^n_t\| := \sup_{i,j\leqslant d} |(Y^n_t)_{i,j}|.$$

Proof. **Step 1**. We notice that on the set $\{\Theta_{t,n} \neq 0\}$ we have $H_l = \overline{H}_l := H_l \mathbb{1}_{\{|Z_l|\leqslant n^{1/4}\}}$. Consequently $J_l = \overline{J}_l := J_l \mathbb{1}_{\{|Z_{l+1}|\leqslant n^{1/4}\}}$ and $\widehat{Y}^n = \overline{Y}^n$ where $(\overline{Y}^n_t)_{t\in\pi_{T,n}}$ is the solution of the equation

$$\overline{Y}^n_{t^n_m} = I - \sum_{l=0}^{m-1} \overline{Y}^n_{t^n_l}(I+\overline{J}_l)^{-1}\overline{J}_l.$$

Moreover, we have

$$\mathbb{E}_{\Theta_{t,n}}[\sup_{s\in\pi_{T,n}} \|\widehat{Y}^n_s\|^p] \leqslant \frac{1}{\mathbb{E}[\Theta_{t,n}]}\mathbb{E}[\sup_{s\in\pi_{T,n}} \|\overline{Y}^n_s\|^p] \leqslant C\mathbb{E}[\sup_{s\in\pi_{T,n}} \|\overline{Y}^n_s\|^p],$$

the last inequality is a consequence of (1.59). Indeed

$$\mathbb{E}[\Theta_{t,n}] \geqslant 1 - \mathbb{P}(\Theta_{t,n}=0) \geqslant 1 - \exp(-m_*^2 nt/(2T)) - \frac{M_8(Z)}{n} \geqslant \frac{1}{2}.$$

The last inequality is true under the hypothesis (1.77). So, our task is now to estimate $\mathbb{E}[\sup_{s\in\pi^T_{T,n}} \|\overline{Y}^n_s\|^p]$.

Step 2. Let

$$\mathcal{F}_l = \sigma(\chi_i, U_i, V_i, i = 1, \ldots, l).$$

Since, from (1.77), the lower eigenvalue of $(I+\overline{J}_l)$ is larger than $1/2$, then $\|(I+\overline{J}_l)^{-1}\| \leqslant 2$. It follows that $\|\overline{Y}^n_{t^n_l}(I+\overline{J}_l)^{-1}\overline{J}_l\| \leqslant 2\|\overline{Y}^n_{t^n_l}\|\|\overline{J}_l\|$ and since $\overline{Y}^n_{t^n_l}$ is $\mathcal{F}_l$ measurable, we obtain

$$\|\mathbb{E}[\overline{Y}^n_{t^n_l}(I+\overline{J}_l)^{-1}\overline{J}_l \mid \mathcal{F}_l]\| \leqslant 2\|\overline{Y}^n_{t^n_l}\|\mathbb{E}[\|\overline{J}_l\| \mid \mathcal{F}_l].$$

Now, we notice that $\mathbb{E}[\|\overline{J}^{p,r}_{l,0}\| \mid \mathcal{F}_l] \leqslant C\|\psi\|_{1,2,\infty}/n$ and

$$\begin{aligned}\mathbb{E}[\|\overline{J}^{p,r}_l(j)\| \mid \mathcal{F}_l] &\leqslant \frac{C}{\sqrt{n}}\|\psi\|_{1,2,\infty}\mathbb{E}[|Z^j_{l+1}|\mathbb{1}_{\{|Z_{l+1}|\geqslant n^{1/4}\}} \mid \mathcal{F}_l] \\ &\leqslant \frac{C}{n}\|\psi\|_{1,2,\infty}\mathbb{E}[|Z_{l+1}|^3] \leqslant \frac{C\|\psi\|_{1,2,\infty}M_3(Z)}{n}.\end{aligned}$$

Moreover, using the Hölder inequality, we obtain

$$\mathbb{E}[\|\overline{J}^{p,r}_l(j,q)\| \mid \mathcal{F}_l]| \leqslant C\frac{M_4(Z)^{1/2}\|\psi\|_{1,3,\infty}}{n}.$$

It follows that $\mathbb{E}[\|\overline{J}_l\| \mid \mathcal{F}_l] \leqslant CM_4(Z)\|\psi\|_{1,3,\infty}/n$ so, finally, we obtain

$$\|\mathbb{E}[\overline{Y}^n_{t^n_l}(I+\overline{J}_l)^{-1}\overline{J}_l \mid \mathcal{F}_l)\| \leqslant CM_4(Z)(1+\|\psi\|_{1,3,\infty})\|\overline{Y}^n_{t^n_l}\|/n. \tag{1.80}$$

Step 3. We are now ready to start our proof. We write

$$(\overline{Y}^n_{t^n_m})^{i,j} = \delta_{i,j} - \sum_{l=0}^{m-1}\theta^{i,j}_l \tag{1.81}$$

with

$$\theta^{i,j}_l = (\overline{Y}^n_{t^n_l}(I+\overline{J}_l)^{-1}\overline{J}_l)^{i,j}.$$

We denote

$$\widehat{\theta}_l = \mathbb{E}[\theta_l \mid \mathcal{F}_l], \qquad \widetilde{\theta}_l = \theta_l - \widehat{\theta}_l$$

and we write

$$\begin{aligned}\overline{Y}^n_{t^n_m} &= M_m + A_m \qquad \text{with} \\ M_m &= -\sum_{l=0}^{m-1}\widetilde{\theta}_l, \qquad A^{i,j}_m = \delta_{i,j} - \sum_{l=0}^{m-1}\widehat{\theta}^{i,j}_l.\end{aligned}$$

By (1.80) we have $\|n\widehat{\theta}_l\| \leqslant CM_4(Z)(1+\|\psi\|_{1,3,\infty})\|\overline{Y}^n_{t^n_l}\|$ and using the triangle inequality, we deduce that

$$\sup_{t^n_k\leqslant t^n_m}\|A_k\| \leqslant 1 + CM_4(Z)(1+\|\psi\|_{1,3,\infty})\frac{1}{n}\sum_{l=0}^{m-1}\|\overline{Y}^n_{t^n_l}\|.$$

So that,

$$\mathbb{E}[\sup_{t^n_k\leqslant t^n_m}\|A_k\|^p]^{1/p} \leqslant 1 + CM_4(Z)(1+\|\psi\|_{1,3,\infty})\frac{1}{n}\sum_{l=0}^{m-1}\|\overline{Y}^n_{t^n_l}\|_p.$$

We recall that $\|\theta_l\| \leqslant 2\|\overline{J}_l\|\|\overline{Y}^n_{t^n_l}\|$ and it follows that

$$\|\widetilde{\theta}_l\| \leqslant \|\theta_l\| + \|\widehat{\theta}_l\| \leqslant C(|Z_{l+1}|^2 + M_4(Z))(1+\|\psi\|_{1,3,\infty})\|\overline{Y}^n_{t^n_l}\|/n^{1/2},$$

and then,

$$\|\widetilde{\theta}_l\|_p^2 \leqslant C(M_{2p}(Z)^{2/p} + M_4(Z)^2)(1+\|\psi\|^2_{1,3,\infty})\|\overline{Y}^n_{t^n_l}\|_p^2/n.$$

Moreover, $(M_m)_{m\in\mathbb{N}^*}$ is a martingale so, using Burkholder's inequality (see (1.102)), we have

$$\mathbb{E}[\sup_{t^n_k\leqslant t^n_m}\|M_k\|^p]^{1/p} \leqslant C(\sum_{l=0}^{m-1}\|\widetilde{\theta}_l\|_p^2)^{1/2}.$$

We conclude that

$$\mathbb{E}[\sup_{t^n_k\leqslant t^n_m}\|\overline{Y}_{t^n_k}\|^p]^{1/p} \leqslant 1 + C(M_{2p}(Z)^{1/p} + M_4(Z))(1+\|\psi\|_{1,3,\infty})(\frac{1}{n}\sum_{l=0}^{m-1}\|\overline{Y}^n_{t^n_l}\|_p^2)^{1/2}.$$

Now, we are going to use the Gronwall's lemma. We put $Q_l = \|\overline{Y}^n_{t^n_l}\|^2$, so that, $\|\overline{Y}^n_{t^n_l}\|_p^2 = \|Q_l\|_{p/2}$. It follows that

$$\mathbb{E}[\sup_{k\leqslant m}Q_k^{p/2}]^{1/p} \leqslant 1 + C(M_{2p}(Z)^{1/p} + M_4(Z))(1+\|\psi\|^2_{1,3,\infty})(\frac{1}{n}\sum_{l=0}^{m-1}\|Q_l\|_{p/2})^{1/2},$$

which gives,

$$\begin{aligned}\|\sup_{k\leqslant m}Q_k\|_{p/2} &\leqslant 1 + C(M_{2p}(Z)^{2/p} + M_4(Z)^2)(1+\|\psi\|^2_{1,3,\infty})\frac{1}{n}\sum_{l=0}^{m-1}\|Q_l\|_{p/2}\\ &\leqslant 1 + C(M_{2p}(Z)^{2/p} + M_4(Z)^2)(1+\|\psi\|^2_{1,3,\infty})\frac{1}{n}\sum_{l=0}^{m-1}\|\sup_{k\leqslant l}Q_k\|_{p/2}.\end{aligned}$$

Then, by Gronwall's lemma,

$$\|\sup_{k\leqslant m}Q_k\|_{p/2} \leqslant \exp\big(C(M_{2p}(Z)^{2/p} + M_4(Z)^2)(1+\|\psi\|^2_{1,3,\infty})\big).$$

The estimate of $\mathbb{E}_{\Theta_{t,n}}[\sup_{s\in\pi_{T,n}}\|Y^n_s\|^p]$ is similar but simpler, so we leave it out.

□

We have the following estimate for the covariance matrix of X^n :

Proposition 1.4.1. *Suppose that there exists $\lambda_* > 0$ such that*

$$\inf_{\kappa\in\mathbb{R}}\inf_{x\in\mathbb{R}^d}\inf_{|\xi|=1}\sum_{i=1}^N\langle\partial_{z_i}\psi(\kappa,x,0,0),\xi\rangle^2 \geqslant \lambda_* \tag{1.82}$$

Assume also that n and $t\in\pi^{0,T}_{T,n}$ are sufficiently large such that (1.77) holds and that

$$n^{1/2} \geqslant \frac{8(N^3+N^2+1)}{\lambda_*}\|\psi\|^2_{1,3,\infty}. \tag{1.83}$$

Let $\sigma_{X^n_t}$ be the Malliavin covariance matrix of X^n_t defined in (1.41) for $t\in\pi_{T,n}$. There exists a constant $C\geqslant 1$, which depends on p, T and the moment of Z up to order $8p$, such that

$$\mathbb{E}_{\Theta_{t,n}}[(\det\sigma_{X^n_t})^{-p}]^{1/p} \leqslant C\frac{\exp(C\|\psi\|^2_{1,3,\infty})}{\lambda_* m_* t/T}. \tag{1.84}$$

.

Proof. Let $t \in \pi^{0,T}_{T,n}$ and $m \in \mathbb{N}^*$ such that $t^n_m = t$. By (1.76), $\sigma_{X^n_t} = Y^n_t \widehat{\sigma}(Y^n_t)^*$, with $(Y^n_t)^*$ the transpose matrix of Y^n_t and $\widehat{\sigma} = \sum_{k=1}^m (\widehat{Y}^n_{t^n_k} I_{k-1}) \times (\widehat{Y}^n_{t^n_k} I_{k-1})^*$. It follows that $\det \sigma_{X^n_t} = (\det Y^n_t)^2 \det \widehat{\sigma}$ andt

$$\mathbb{E}_{\Theta_{t,n}}[(\det \sigma_{X^n_t})^{-p}] \leqslant \mathbb{E}_{\Theta_{t,n}}[(\det Y^n_t)^{-4p}]^{1/2} \mathbb{E}_{\Theta_{t,n}}[(\det \widehat{\sigma})^{-2p}]^{1/2}.$$

Since $(\det Y^n_t)^{-1} = \det \widehat{Y}^n_t$, we use (1.79) and we obtain $\mathbb{E}_{\Theta_{t,n}}[(\det Y^n_t)^{-4p}]^{1/2} \leqslant \exp(C(M_{8p}(Z)^{1/(2p)} + M_4(Z)^2)(1 + \|\psi\|^2_{1,3,\infty}))$. We estimate now the lower eigenvalue of $\widehat{\sigma}$ given by

$$\widehat{\lambda} = \inf_{|\xi|=1} \sum_{k=1}^m \sum_{i=1}^N \left\langle (\widehat{Y}^n_{t^n_k} I_{k-1,i}) \times (\widehat{Y}^n_{t^n_k} I_{k-1,i})^* \xi, \xi \right\rangle = \inf_{|\xi|=1} \sum_{k=1}^m \sum_{i=1}^N \left\langle (I_{k-1,i} I_{k-1,i})^* (\widehat{Y}^n_{t^n_k})^* \xi, (\widehat{Y}^n_{t^n_k})^* \xi \right\rangle. \tag{1.85}$$

Recall that, $I_{k,i}$ is given in (1.74):

$$I_{k,i} = \frac{1}{\sqrt{n}} \chi_{k+1} (a^i_k + \sum_{j=1}^N H^j_{k+1} b^{i,j}_k + \frac{1}{\sqrt{n}} \sum_{j,q=1}^N H^j_{k+1} H^q_{k+1} c^{i,j,q}_k + \delta^n_{k+1} \tilde{c}^i_k).$$

Then, for $\eta \in \mathbb{R}^d$ and $k \in \{0, \ldots, M-1\}$ we have

$$\begin{aligned}
\sum_{i=1}^N \langle (I_{k,i} I_{k,i})^* \eta, \eta \rangle &= \sum_{i=1}^N \langle I_{k,i}, \eta \rangle^2 \\
&\geqslant \frac{1}{2n} \sum_{i=1}^N \chi_{k+1} \left\langle a^i_k, \eta \right\rangle^2 \\
&- 2(N^3 + N^2 + 1) \sup_{i,j,q} \{ \, | \left\langle H^j_{k+1} b^{i,j}_k, \eta \right\rangle |^2, | \left\langle H^j_{k+1} H^q_{k+1} c^{i,j,q}_k, \eta \right\rangle |^2, |\langle \delta^n_{k+1} \tilde{c}^i_k, \eta \rangle|^2 \}.
\end{aligned}$$

Since we are on the set $\{\Theta_{t,n} \neq 0\}$, we have $\sup_{k \in \{1,..,n\}} |H_k| \leqslant n^{-1/4}$. Moreover, $\sup_{i,j,q} \{|b^{i,j}_k|, |c^{i,j,q}_k|, |\tilde{c}^i_k|\} \leqslant \|\psi\|_{1,3,\infty}$, for all $k \in \{0, \ldots, n-1\}$, so that

$$\sup_{i,j,q} \{ \, | \left\langle H^j_{k+1} b^{i,j}_k, \eta \right\rangle |, | \left\langle H^j_{k+1} H^q_{k+1} c^{i,j,q}_k, \eta \right\rangle |, |\langle \delta^n_{k+1} \tilde{c}^i_k, \eta \rangle| \} \leqslant \frac{1}{n^{1/4}} \|\psi\|_{1,3,\infty} |\eta|.$$

We recall that we have the hypothesis (1.82)

$$\sum_{i=1}^N \left\langle a^i_k, \eta \right\rangle^2 = \sum_{i=1}^N \left\langle \partial_{z_i} \psi_k(X^n_{t^n_k}, 0, 0), \eta \right\rangle^2 \geqslant \lambda_* |\eta|^2.$$

Using (1.83), we have $\lambda_*/2 - 2(N^3 + N^2 + 1)\|\psi\|^2_{1,3,\infty}/n^{1/2} \geqslant \lambda_*/4$, and we obtain

$$\sum_{i=1}^N \langle (I_{k,i} I_{k,i})^* \eta, \eta \rangle \geqslant \frac{\chi_{k+1}}{n} (\frac{\lambda_*}{2} - 2 \frac{(N^3 + N^2 + 1)\|\psi\|^2_{1,3,\infty}}{n^{1/2}}) |\eta|^2 \geqslant \chi_{k+1} \frac{1}{4n} \lambda_* |\eta|^2$$

We come back to (1.85) and we take $\eta = (\widehat{Y}^n_{t^n_k})^* \xi$. Since on the set $\{\Theta_{t,n} \neq 0\}$ and $\lfloor nt/T \rfloor = nt/T$, we have $\frac{T}{nt} \sum_{k=1}^{nt/T} \chi_k \geqslant \frac{1}{2} m_*$, it follows that

$$\begin{aligned}
\widehat{\lambda} &\geqslant \frac{\lambda_*}{4} \frac{1}{n} \inf_{|\xi|=1} \sum_{k=1}^{nt/T} \chi_k \|(\widehat{Y}^n_{t^n_k})^* \xi\|^2 \geqslant \frac{\lambda_*}{4n} \sum_{k=1}^{nt/T} \chi_k \inf_{|\xi|=1} \|(\widehat{Y}^n_{t^n_k})^* \xi\|^2 \\
&\geqslant \frac{\lambda_* m_* nt/T}{8n} \inf_{s \in \pi^T_{T,n}; s \leqslant t} \inf_{|\xi|=1} \|(\widehat{Y}^n_s)^* \xi\|^2 \geqslant \frac{\lambda_* m_* t}{8T} (\sup_{s \in \pi^T_{T,n}; s \leqslant t} \|Y^n_s\|)^{-2}
\end{aligned}$$

Since we have (1.77), (1.78) follows and we conclude that

$$\begin{aligned}\mathbb{E}_{\Theta_{t,n}}[\widehat{\lambda}^{-p}]^{1/p} \leqslant & \frac{8n}{\lambda_* m_*(nt/T)}\mathbb{E}_{\Theta_{t,n}}[\sup_{s\in\pi^T_{T,n}} \|Y^n_s\|^{2p}]^{1/p} \\ \leqslant & C\frac{\exp\left(C(M_{4p}(Z)^{1/p}+M_4(Z)^2)(1+\|\psi\|^2_{1,3,\infty})\right)T}{\lambda_* m_* t}.\end{aligned}$$

□

1.4.2 The regularization property

We still fix $T>0$ and $n\in\mathbb{N}^*$ and we consider the Markov chain $(X^n_t)_{t\in\pi_{T,n}}$, defined in (1.65). We also recall that $\Theta_{S,n}$ is defined in (1.58) and we introduce $(Q^{n,\Theta}_t)_{t\in\pi_{T,n}}$ such that,

$$\forall t\in\pi_{T,n},\quad Q^{n,\Theta}_t f(x):=\mathbb{E}_{\Theta_{t,n}}[f(X^n_t(x))]=\frac{1}{\mathbb{E}[\Theta_{t,n}]}\mathbb{E}[\Theta_{t,n}f(X^n_t(x))]. \tag{1.86}$$

Notice that $(Q^{n,\Theta}_t)_{t\in\pi_{T,n}}$, is not a semigroup, but this is not necessary. We will not be able to prove the regularization property for Q^n but for $Q^{n,\Theta}$ and every $t\leqslant T$.

Proposition 1.4.2. ***A.*** *Let $T>0$ and $n\in\mathbb{N}^*$. We assume that n and $t\in\pi^{0,T}_{T,n}$ are sufficiently large in order to have (1.77) :*

$$\frac{3\|\psi\|_{1,3,\infty}}{n^{1/4}}+\frac{M_8(Z)}{n}+\exp(-m_*^2 nt/(2T))\leqslant\frac{1}{2}$$

and (1.83). Moreover we assume that (1.82) holds true. Then for every $q\in\mathbb{N}$ and multi index α,β with $|\alpha|+|\beta|\leqslant q$, there exists $l\in\mathbb{N}^*$ and $C\geqslant 1$ which depend on m_*,r_* and the moments of Z such that

$$\|\partial_\alpha Q^{n,\Theta}_t\partial_\beta f\|_\infty \leqslant C\frac{\mathfrak{K}_{q+3}(\psi)^l}{(\lambda_* t)^{q(q+1)}}\|f\|_\infty \tag{1.87}$$

with $\mathfrak{K}_r(\psi)$ defined in (1.69). In particular, $Q^{n,\Theta}_t(x,dy)=p^{n,\Theta}_t(x,y)dy$ and $(x,y)\mapsto p^{n,\Theta}_t(x,y)$ belongs to $\mathcal{C}^\infty(\mathbb{R}^d\times\mathbb{R}^d)$.

B. *There exists $C\geqslant 1$, such that for every $l\in\mathbb{N}$ and $t\in\pi^T_{T,n}$, we have*

$$\|Q^n_t f-Q^{n,\Theta}_t f\|_\infty\leqslant 4(\exp(-m_*^2nt/(2T))+\frac{M_{4(l+1)}(Z)}{n^l})\|f\|_\infty. \tag{1.88}$$

Remark 1.4.2. *(1.87) means that the strong regularization property $\overline{R}_{q,\eta}$ (see (1.25)), with $\eta(q)=q(q+1)$, holds for $Q^{n,\Theta}$.*

Proof. We fix $t\in\pi^{0,T}_{T,n}$. Let us prove **A**.

$$\partial_\alpha Q^{n,\Theta}_t\partial_\beta f(x)=\sum_{|\beta|\leqslant|\gamma|\leqslant q}\mathbb{E}_{\Theta_{t,n}}[\partial_\gamma f(X^n_t(x))\mathcal{P}_\gamma(X^n_t)], \tag{1.89}$$

where $\mathcal{P}_\gamma(X^n_t)$ is a universal polynomial of $\partial^\rho_x X^n_t(x), 0\leqslant|\rho|\leqslant q-|\gamma|+1$. Using the integration by parts formula (1.61) and the estimate (1.63) (together with $\mathbb{E}[\Theta_{t,n}]\geqslant 1/2$ using (1.77)) we obtain

$$\begin{aligned}|\mathbb{E}_{\Theta_{t,n}}[\partial_\gamma f(X^n_t(x))\mathcal{P}_\gamma(X^n_t(x))]| &= |\mathbb{E}_{\Theta_{t,n}}[f(X^n_t(x))H^{\Theta_{t,n}}_\gamma(X^n_t(x),\mathcal{P}_\gamma(X^n_t(x))]| \\ &\leqslant \|f\|_\infty\mathbb{E}_{\Theta_{t,n}}[|H^{\Theta_{t,n}}_\gamma(X^n_t(x),\mathcal{P}_\gamma(X^n_t(x))|] \\ &\leqslant C\|f\|_\infty\times A_1\times A_2\times A_3\end{aligned} \tag{1.90}$$

with

$$
\begin{aligned}
A_1 &= 1 \vee \mathbb{E}_{\Theta_{t,n}}[((\det \sigma_{X_t^n(x)})^{-1})^{2q(q+1)}]^{1/2} \\
A_2 &= 1 + \mathbb{E}[|X_t^n(x)|_{1,q+1}^{8qd(q+2)}]^{1/4} + \mathbb{E}[|LX_t^n(x)|_{q-1}^{8q}]^{1/4} \\
A_3 &= \mathbb{E}[|\mathcal{P}_\gamma(X_t^n(x))|_{|\gamma|}^4]^{1/4}.
\end{aligned}
$$

Using the results from Theorem 1.4.1, we obtain

$$A_2 \times A_3 \leqslant C\mathfrak{K}_{q+3}(\psi)^l.$$

We use now (1.84) and it follows

$$A_1 = 1 \vee \mathbb{E}_{\Theta_{t,n}}[(\det \sigma_{X_t^n(x)})^{-2q(q+1)}]^{1/2} \leqslant 1 \vee C(\lambda_* t)^{-q(q+1)} \exp(Cq(q+1)\|\psi\|_{1,3,\infty}^2).$$

Now, we gather all the terms together,

$$|\partial_\alpha Q_t^{n,\Theta} \partial_\beta f(x))| \leqslant C \frac{\mathfrak{K}_{q+3}(\psi)^l}{(\lambda_* t)^{q(q+1)}} \|f\|_\infty.$$

B. We have

$$
\begin{aligned}
|Q_t^n f(x) - Q_t^{n,\Theta} f(x)| &\leqslant |Q_t^n f(x)||1 - \frac{1}{\mathbb{E}[\Theta_{t,n}]}| + \frac{1}{\mathbb{E}[\Theta_{t,n}]}|\mathbb{E}[f(X_t^n(x))(1-\Theta_{t,n})]| \\
&\leqslant 2\|f\|_\infty \frac{\mathbb{E}[|1-\Theta_{t,n}|]}{\mathbb{E}[\Theta_{t,n}]} \leqslant 2\|f\|_\infty \frac{\mathbb{P}(\Theta_{t,n}=0)}{1-\mathbb{P}(\Theta_{t,n}=0)}.
\end{aligned}
$$

By (1.59) we have, for every $l \in \mathbb{N}$, $\mathbb{P}(\Theta_{t,n} = 0) \leqslant \exp(-m_*^2 nt/(2T)) + M_{4(l+1)}(Z)n^{-l}$ and we conclude using (1.77) in order to obtain $1 - \mathbb{P}(\Theta_{t,n} = 0) \geqslant 1/2$. □

We give now an alternative way to regularize the semigroup Q^n (by convolution). We consider a d dimensional standard normal random variable G which is independent from $Z_k, k \in \mathbb{N}^*$, and for $\theta > 0$, we introduce $(X_t^{n,\theta})_{t\in\pi_{T,n}}$ as follows

$$X_t^{n,\theta}(x) = \frac{1}{n^\theta} G + X_t^n(x). \tag{1.91}$$

We denote by $p_t^{n,\theta}(x,y)$ the density of the law of $X_t^{n,\theta}(x)$ and for $t \in \pi_{T,n}$, we define

$$Q_t^{n,\theta} f(x) := \mathbb{E}[f(\frac{1}{n^\theta} G + X_t^n(x))]. \tag{1.92}$$

Corollary 1.4.1. *Under the hypothesis of the previous proposition we have:*

A. *For every multi index α, β with $|\alpha| + |\beta| \leqslant q$, and every $q \in \mathbb{N}^*$, there exists $l \in \mathbb{N}^*$, $C \geqslant 1$, which depend on q, T and the moments of Z such that for all $l' \in \mathbb{N}$ and $t \in \pi_{T,n}^{0,T}$ sufficiently large in order to have (1.77) and (1.83), the following estimate holds :*

$$\|\partial_\alpha Q_t^{n,\theta} \partial_\beta f\|_\infty \leqslant C\Big(\frac{\mathfrak{K}_{q+3}(\psi)^l}{(\lambda_* t)^{q(q+1)}} + n^{q\theta}\mathfrak{K}_{q+3}(\psi)^l(\exp(-m_*^2 nt/(4T)) + \frac{M_{4(l'+1)}(Z)^{1/2}}{n^{l'/2}})\Big)\|f\|_\infty, \tag{1.93}$$

with $\mathfrak{K}_r(\psi)$ defined in (1.87).

B. *There exists $l \in \mathbb{N}^*$, $C \geqslant 1$, such that for every $l' \in \mathbb{N}$ and $t \in \pi^{T}_{T,n}$*

$$\|Q^n_t f(x) - Q^{n,\theta}_t f(x)\|_\infty \leqslant C\Big(\frac{\mathfrak{K}_4(\psi)^l}{(\lambda_* t)^2 n^\theta} + 2(\exp(-m_*^2 nt/(2T)) + \frac{M_{4(l'+1)}(Z)}{n^{l'}}\Big)\Big)\|f\|_\infty. \quad (1.94)$$

Proof. We fix $t \in \pi^{0,T}_{T,n}$. Let us prove **A**. As in (1.89), we write

$$\partial_\alpha Q^{n,\theta}_t \partial_\beta f(x) = \sum_{|\beta|\leqslant|\gamma|\leqslant q} \mathbb{E}[(\partial_\gamma f)(n^{-\theta}G + X^n_t(x))\mathcal{P}_\gamma(X^n_t(x))],$$

where $\mathcal{P}_\gamma(X^n_t)$ is a universal polynomial of $\partial^\rho_x X^n_t(x), 0 \leqslant |\rho| \leqslant q - |\gamma| + 1$. We decompose

$$\mathbb{E}[(\partial_\gamma f)(n^{-\theta}G + X^n_t(x))\mathcal{P}_\gamma(X^n_t(x))] = I + J$$

with

$$\begin{aligned} I &= \mathbb{E}[\Theta_{t,n}]\mathbb{E}_{\Theta_{t,n}}[\partial_\gamma f(n^{-\theta}G + X^n_t(x))\mathcal{P}_\gamma(X^n_t(x))], \\ J &= \mathbb{E}[(\partial_\gamma f)(n^{-\theta}G + X^n_t(x))\mathcal{P}_\gamma(X^n_t(x))(1 - \Theta_{t,n})]. \end{aligned}$$

The reasoning from the previous proof shows that

$$I \leqslant C\frac{\mathfrak{K}_{q+3}(\psi)^l}{(\lambda_* t)^{q(q+1)}}\|f\|_\infty.$$

And since G follows the standard normal law and is independent from X^n and $\Theta_{t,n}$, we have

$$J = \mathbb{E}\big[\mathcal{P}_\gamma(X^n_t(x))(1 - \Theta_{t,n}) \int_{\mathbb{R}^d} (\partial_\gamma f)(n^{-\theta}y + X^n_t(x))(2\pi)^{-d/2}e^{-|y|^2/2}dy\big].$$

Moreover, one has

$$(\partial_\gamma f)(n^{-\theta}y + X^n_t(x)) = n^{|\gamma|\theta}\partial^\gamma_y(f(n^{-\theta}y + X^n_t(x))),$$

so that, using standard integration by parts, we have

$$J = n^{|\gamma|\theta}\mathbb{E}\big[\mathcal{P}_\gamma(X^n_t(x))(1 - \Theta_{t,n}) \int_{\mathbb{R}^d} f(n^{-\theta}y + X^n_t(x))H_\gamma(y)(2\pi)^{-d/2}e^{-|y|^2/2}dy\big],$$

where H_γ is the Hermite polynomial corresponding to the multi-index γ. Finally we obtain

$$|J| \leqslant Cn^{|\gamma|\theta}\mathfrak{K}_{q+3}(\psi)^l\|f\|_\infty \mathbb{E}[1 - \Theta_{t,n}]^{1/2} \leqslant Cn^{|\gamma|\theta}\mathfrak{K}_{q+3}(\psi)^l\|f\|_\infty(\exp(-m_*^2 nt/(4T)) + \frac{M_{4(l'+1)}(Z)^{1/2}}{n^{l'/2}})$$

the last inequality being a consequence of (1.59).
Now we prove **B**. Let $l' \in \mathbb{N}^*$. Using (1.59) and (1.87), there exists $C, l \geqslant 1$ such that we have

$$\begin{aligned} |Q^n_t f(x) - Q^{n,\theta}_t f(x)| &\leqslant \mathbb{E}[\Theta_{t,n}]|\mathbb{E}_{\Theta_{t,n}}[f(X^n_t(x)) - f(X^n_t(x) + n^{-\theta}G)]| + 2\|f\|_\infty \mathbb{E}[1 - \Theta_{t,n}] \\ &\leqslant n^{-\theta}\sum_{j=1}^d \int_0^1 |\mathbb{E}_{\Theta_{t,n}}[\partial_j f(X^n_t(x) + \lambda n^{-\theta}G)G^j]|d\lambda + 2\|f\|_\infty \mathbb{E}[1 - \Theta_{t,n}] \\ &\leqslant Cn^{-\theta}\frac{\mathfrak{K}_4(\psi)^l}{(\lambda_* t)^2}\|f\|_\infty + 2(\exp(-m_*^2 nt/(2T)) + \frac{M_{4(l'+1)}(Z)}{n^{l'}})\|f\|_\infty. \end{aligned}$$

□

1.4.3 Approximation result

In this section we give the approximation result for a Markov semigroup $(P_t)_{t\geqslant 0}$. We recall that $T>0$ and $n\in\mathbb{N}$ are fixed. We denote $\mu_k^n(x,dy)=P_{T/n}(x,dy)$ for all $k\in\mathbb{N}^*$. We consider now an approximation scheme based on the Markov chain introduced in the previous section (see (1.65). Therefore, we consider two sequences of independent random variables $Z_{k+1}\in\mathbb{R}^N$, $\kappa_k\in\mathbb{R}$, $k\in\mathbb{N}$ and we take $(\delta_k^n)_{k\in\mathbb{N}^*}$ such that $\sup_{k\in\mathbb{N}^*}\delta_k^n\leqslant C/n$ for a constant $C\geqslant 1$. We assume that $Z_1,\ldots,Z_n$ verifies (1.32) and have finite moments of any order: For every $p\geqslant 1$,

$$M_p(Z)=1\vee\sup_{k\leqslant n}\mathbb{E}[|Z_k|^p]<\infty. \tag{1.95}$$

Moreover, we take $\psi\in\mathcal{C}^\infty(\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^N\times\mathbb{R}_+;\mathbb{R}^d)$ such that $\psi(\kappa,x,0,0)=x$ and we construct $X^n_{t^n_{k+1}}(x)=\psi(\kappa_k,X^n_{t^n_k}(x),Z_{k+1}/\sqrt{n},\delta^n_{k+1})$ with $X^n_0(x)=x$. We denote $\nu^n_{k+1}(x,dy)=\mathbb{P}(X^n_{t^n_{k+1}}\in dy\mid X^n_{t^n_k}=x)$ and we construct the discrete semigroup $Q^n_{t^n_{k+1}}=Q^n_{t^n_k}\nu^n_{k+1}$ on the time grid $\pi_{T,n}$. We recall that the notation $\|\psi\|_{1,r,\infty}$ is introduced in (1.67) and we assume that, for every $r\in\mathbb{N}$,

$$\|\psi\|_{1,r,\infty}<\infty. \tag{1.96}$$

We also assume that there exists $\lambda_*>0$ such that

$$\inf_{\kappa\in R}\inf_{x\in\mathbb{R}^d}\inf_{|\xi|=1}\sum_{i=1}^N\langle\partial_{z_i}\psi(\kappa,x,0,0),\xi\rangle^2\geqslant\lambda_*. \tag{1.97}$$

Now we are able to prove our main result.

Theorem 1.4.2. *We recall that $T>0$. We fix $q\in\mathbb{N}$, $h>0$ and $S\in(0,T/2)$. For a given $n\in\mathbb{N}^*$, we consider the Markov semigroup $(P_t)_{t\geqslant 0}$, and the approximation Markov chain $(Q_t^n)_{t\in\pi_{T,n}}$, defined above. Moreover, we assume that there exists $n_0\in\mathbb{N}^*$ such that $T/n_0\leqslant S$ and, (1.77) and (1.83) hold with $n=n_0$ and $t=S$. Then, for all $n\geqslant n_0$, we have the following properties.*

- ***A.*** *We assume that (1.95), (1.96) and (1.97) hold. Moreover we assume that $E_m(h,q)$ (see (1.16)) and $E^*_m(h,q)$ (see (1.19)) hold between $(P^m_t)_{t\in\pi_{T,m}}=(P_t)_{t\in\pi_{T,m}}$ and $(Q^m_t)_{t\in\pi_{T,m}}$ for every $m\geqslant n$. Then, there exists $l\in\mathbb{N}^*$, $C\geqslant 1$, which depend on q, T and the moments of Z, such that*

$$\sup_{t\in\pi^{2S,T}_{T,n}}\|P_tf-Q^n_tf\|_\infty\leqslant C\frac{\mathfrak{K}_{q+3}(\psi)^l}{(\lambda_*S)^{\eta(q)}}\|f\|_\infty/n^h. \tag{1.98}$$

  with $\eta(q)=q(q+1)$.

- ***B.*** *Moreover, for every $t>0, P_t(x,dy)=p_t(x,y)dy$ with $(x,y)\mapsto p_t(x,y)$ belonging to $\mathcal{C}^\infty(\mathbb{R}^d\times\mathbb{R}^d)$.*

- ***C.*** *We recall the $Q^{n,\Theta}$ is defined in (1.86) and verifies $Q^{n,\Theta}_t(x,dy)=p^{n,\Theta}_t(x,y)dy$. Then, there exists $l\in\mathbb{N}^*$ such that for every $R>0,\varepsilon\in(0,1)$, $x_0,y_0\in\mathbb{R}^d$, and every multi-index α, β with $|\alpha|+|\beta|=u$, we also have*

$$\sup_{t\in\pi^{2S,T}_{T,n}}\sup_{(x,y)\in\overline{B}_R(x_0,y_0)}|\partial_x^\alpha\partial_y^\beta p_t(x,y)-\partial_x^\alpha\partial_y^\beta p^{n,\Theta}_t(x,y)|\leqslant C\frac{\mathfrak{K}_{q+3}(\psi)^l}{(\lambda_*S)^{\eta((p_{u,\varepsilon}\vee q)}}/n^{h(1-\varepsilon)} \tag{1.99}$$

  with a constant C which depends on R,x_0,y_0,T and on $|\alpha|+|\beta|$ and $p_{u,\varepsilon}=(u+2d+1+2\lceil(1-\varepsilon)(u+d)/(2\varepsilon)\rceil)$.

D. *Let $\theta \geqslant h+1$. We recall the $Q^{n,\theta}$ is defined in (1.92) and verifies $Q_t^{n,\theta}(x,dy) = p_t^{n,\theta}(x,y)dy$. Then, there exists $l \in \mathbb{N}^*$ such that for every $R > 0, \varepsilon \in (0,1)$, $x_0, y_0 \in \mathbb{R}^d$, and every multi-index α, β with $|\alpha| + |\beta| = u$, we also have*

$$\sup_{t \in \pi_{T,n}^{2S,T}} \sup_{(x,y)\in \overline{B}_R(x_0,y_0)} |\partial_x^\alpha \partial_y^\beta p_t(x,y) - \partial_x^\alpha \partial_y^\beta p_t^{n,\theta}(x,y)| \leqslant C \frac{\mathfrak{K}_{q+3}(\psi)^l}{(\lambda_* S)^{\eta(p_{u,\varepsilon}\vee q)}} / n^{h(1-\varepsilon)} \tag{1.100}$$

Proof. **A-B**. We use Proposition 1.2.3: We have proved in Proposition 1.4.2 that $Q^{n,\Theta}$ verifies the regularization properties. The proof of (1.98) and (1.99) is an immediate consequence of Theorem 1.2.1. **C**. In order prove (2.24) one uses Corollary 1.4.1 instead of Proposition 1.4.2. □

Remark 1.4.3. *The simulation of an approximation scheme given by $Q^{n,\Theta}$ may be cumbersome, so the estimate obtained in (1.99) is not very useful. This is why we propose the regularized scheme $X^{n,\theta}$ which is easier to simulate.*

1.4.4 Proof of Theorem 1.4.1 on Sobolev norms

In this section, we will obtain estimates of the Sobolev norms of X^n and LX^n which appear in Theorem 1.4.1. The method we adopt here is to prove the estimates for a generic class of processes which involves the Malliavin derivatves of X^n and LX^n.

Before doing it, we give some preliminary results. We consider a separable Hilbert space U, we denote $|a|_U$ the norm of U and, for a random variable $F \in U$, we denote $\|F\|_{U,p} = (\mathbb{E}[|F|_U^p)]^{1/p}$. Moreover we consider a martingale $M_n \in U$, $n \in \mathbb{N}$ and we recall Burkholder's inequality in this framework: For each $p \geqslant 2$ there exists a constant $b_p \geqslant 1$ such that

$$\forall n \in \mathbb{N}, \quad \|M_n\|_{U,p} \leqslant b_p \mathbb{E}[(\sum_{k=1}^n |M_k - M_{k-1}|_U^2)^{p/2}]^{1/p}. \tag{1.101}$$

As an immediate consequence

$$\|M_n\|_{U,p} \leqslant b_p (\sum_{k=1}^n \|M_k - M_{k-1}\|_{U,p}^2)^{1/2}. \tag{1.102}$$

Indeed

$$\begin{aligned} \|M_n\|_{U,p}^2 &\leqslant b_p^2 \mathbb{E}[(\sum_{k=1}^n |M_k - M_{k-1}|_U^2)^{p/2}]^{2/p} = b_p^2 \| \sum_{k=1}^n |M_k - M_{k-1}|_U^2 \|_{p/2} \\ &\leqslant b_p^2 \sum_{k=1}^n \| |M_k - M_{k-1}|_U^2 \|_{p/2} = b_p^2 \sum_{k=1}^n \|M_k - M_{k-1}\|_{U,p}^2. \end{aligned}$$

We consider the scheme defined in the previous sections (see (1.70)) :

$$X_{t_{k+1}^n}^n = x + \sum_{i=1}^N \sum_{k=0}^{m-1} H_{k+1}^i a_k^i (X_{t_k^n}^n) + \sum_{k=0}^{m-1} \delta_{k+1}^n \tilde{b}_k (X_{t_k^n}^n, H_{k+1}, \delta_{k+1}^n) + \frac{1}{2} \sum_{i,j=1}^N \sum_{k=0}^{m-1} H_{k+1}^i H_{k+1}^j b_k^{i,j} (X_{t_k^n}^n, H_{k+1})$$

with $H_k = n^{-1/2} Z_k$ and

$$a_k^i(x) = \partial_{z_i}\psi(\kappa_k, x, 0, 0),\ b_k^{i,j}(x,z) = \int_0^1 (1-\lambda)\partial_{z_i}\partial_{z_j}\psi(\kappa_k, x, \lambda z, 0)d\lambda,\ \tilde{b}_k(x,z,t) = \int_0^1 \partial_t \psi(\kappa_k, x, z, \lambda t)d\lambda.$$

We also denote

$$A_k = \sum_{i=1}^{N} H^i_{k+1}\nabla_x a^i_k(X^n_{t^n_k}) + \delta^n_{k+1}\nabla_x \tilde{b}_k(X^n_{t^n_k}, H_{k+1}, \delta^n_{k+1}) + \frac{1}{2}\sum_{i,j=1}^{N} H^i_{k+1}H^j_{k+1}\nabla_x b^{i,j}_k(X^n_{t^n_k}, H_{k+1}).$$

Notice that $X^n_t, a^i_k, b^{i,j}_k, \tilde{b}_k \in \mathbb{R}^d$ and A_k is a $d \times d$ dimensional matrix.

Now, we focus on the estimates of the Sobolev norms. As before, U is a separable Hilbert space. We say that, a U valued random variable F belongs to $\mathcal{S}(U)$ if for every $h \in U$ we have $\langle h, F\rangle \in \mathcal{S}$ (see (1.39)) and we define DF by $\langle h, DF\rangle = D\langle h, F\rangle$ for every $h \in U$. Then, we define the norms (see (1.51) and (1.52))

$$|F|^2_{U,m} = \sum_{0\leqslant|\alpha|\leqslant m} |D_\alpha F|^2_U, \qquad \|F\|_{U,m,p} = \left\||F|_{U,m}\right\|_p = \mathbb{E}[|F|^p_{U,m}]^{1/p}.$$

The Hilbert space U being given, we denote $V = U^d$ (recall that $X^n_{t^n_k} \in \mathbb{R}^d$ so, in this case, $U = \mathbb{R}$ and $V = \mathbb{R}^d$). We consider now some processes $(\alpha_k)_{k\in\mathbb{N}}, (\beta_k)_{k\in\mathbb{N}}, (\Gamma_k)_{k\in\mathbb{N}}$ with $\alpha_k = (\alpha^1_k, \ldots, \alpha^N_k) \in V^N$, $\beta_k = (\beta^1_k, \ldots, \beta^N_k) \in V^N$, $\Gamma_k \in V$. We assume that $\alpha^i_k = \alpha^i_k(Z_1, \ldots, Z_k)$ and $\langle h, \alpha^i_k\rangle \in \mathcal{C}^\infty_b(\mathbb{R}^{kN})$ for every $h \in V, i = 1, \ldots, N$ (we recall that $Z_k \in \mathbb{R}^N$). So $\alpha_k \in \mathcal{S}(V)$. The same is assumed on β_k and Γ_k. We look at a process $Y_k \in V = U^d$, $k \in \mathbb{N}$ which satisfies the equation

$$Y_m = Y_0 + \sum_{k=0}^{m-1} A_k Y_k + \sum_{i=1}^{N}\sum_{k=0}^{m-1} H^i_{k+1}\alpha^i_k + \sum_{i=1}^{N}\sum_{k=0}^{m-1} LH^i_{k+1}\beta^i_k + \Gamma_m. \tag{1.103}$$

Notice that we do not discuss about existence and uniqueness of the solution of such an equation. We just suppose that, the process Y at hand satisfies this equation (which naturally appears in our calculus). We aim to estimate the Sobolev norms of Y_m. Let $q \in \mathbb{N}$ and $p \geqslant 2$. We denote

$$C_{q,p}(\alpha, \beta, \Gamma) = \sup_{0\leqslant m\leqslant n-1}\sup_{i=1,..,N}(1 + \|\alpha^i_m\|_{V,q,p} + \|\beta^i_m\|_{V,q,p} + \|\Gamma_{m+1}\|_{V,q,p}) \tag{1.104}$$

Proposition 1.4.3. *For every $q \in \mathbb{N}$ and $p \geqslant 2$ there exists some constants $l \in \mathbb{N}^*$, $C \geqslant 1$ (depending on q and p) such that*

$$\sup_{m\leqslant n}\|Y_m\|_{V,q,p} \leqslant C(M_l(Z) + \frac{m_*^{1/l}}{r_*}(1 + r_*^{-q}))C_{q,l}(\alpha, \beta, \Gamma)\mathfrak{K}_{q+2}(CM_l(Z)\psi)^l. \tag{1.105}$$

with $\mathfrak{K}_r(\psi)$ and $M_l(Z)$ defined in (1.69) and (1.33).

Proof. **Step 1**. Let $q = 0$, so that $\|Y_m\|_{V,q,p} = \|Y_m\|_{V,p}$. We will check that

$$\sup_{m\leqslant n}\|Y_m\|_{V,p} \leqslant C\big(M_p(Z)^{1/p}C_{0,p}(\alpha, 0, 0) + \frac{m_*^{1/p}}{r_*}C_{0,p}(0, \beta, 0) + C_{0,p}(0, 0, \Gamma)\big)$$
$$\times \exp(CM_{2p}(Z)^{2/p}\|\psi\|^2_{1,3,\infty}). \tag{1.106}$$

We study the terms which appear in the right hand side of (1.103). Notice that β^i_k is $\sigma(Z_1, \ldots, Z_k)$ measurable and $\mathbb{E}[LH^i_{k+1}] = 0$ (see (1.54)). It follows that, $M_m = \sum_{k=0}^{m-1} LH^i_{k+1}\beta^i_k$ is a martingale and consequently, by (1.102)

$$\|M_m\|_{V,p} \leqslant b_p(\sum_{k=0}^{m-1}\|LH^i_{k+1}\beta^i_k\|^2_{V,p})^{1/2}.$$

Since LH^i_{k+1} and β^i_k are independent, using (1.55) we obtain

$$\|LH^i_{k+1}\beta^i_k\|^2_{V,p} = \|LH^i_{k+1}\|^2_p\|\beta^i_k\|^2_{V,p} \leqslant \frac{Cm_*^{2/p}}{r_*^2}\|\beta^i_k\|^2_{V,p}/n.$$

We conclude that

$$\sup_{m\leqslant n}\|M_m\|_{V,p} \leqslant \frac{Cm_*^{1/p}}{r_*}(\frac{1}{n}\sum_{k=0}^{n-1}\|\beta^i_k\|^2_{V,p})^{1/2} \leqslant \frac{Cm_*^{1/p}}{r_*}\sup_{k\leqslant n-1}\|\beta^i_k\|_{V,p}.$$

Since H^i_{k+1} is independent from α^i_k and $\mathbb{E}[H^i_{k+1}] = 0$, it follows that $M_m = \sum_{k=0}^{m-1} H^i_{k+1}\alpha^i_k$ is a martingale. We have $\|H^i_k\|_p \leqslant n^{-1/2}M_p(Z)^{1/p}$ so the same reasoning as above proves that the previous inequality holds for M_m (with $m_*^{1/p}r_*^{-1}$ replaced by $M_p(Z)^{1/p}$ and $\|\beta^i_k\|_{V,p}$ replaced by $\|\alpha^i_k\|_{V,p}$). We use the same reasoning for $M_m = \sum_{k=0}^{m-1} H^i_{k+1}\nabla_x a^i_k(X^n_{t^n_k})Y_k \in V$ and we obtain

$$\|M_m\|_{V,p} \leqslant b_p(\sum_{k=0}^{m-1}\|H^i_{k+1}\nabla_x a^i_k(X^n_{t^n_k})Y_k\|^2_{V,p})^{1/2} \leqslant CM_p(Z)^{1/p}\|\psi\|_{1,2,\infty}(\frac{1}{n}\sum_{k=0}^{m-1}\|Y_k\|^2_{V,p})^{1/2}.$$

Finally, using the triangle inequality

$$\begin{aligned}
\|\sum_{k=0}^{m-1} H^i_{k+1}H^j_{k+1}\nabla_x b^{i,j}_k(X^n_{t^n_k}, H_{k+1})Y_k\|_{V,p} &\leqslant \sum_{k=0}^{m-1}\|H^i_{k+1}H^j_{k+1}\nabla_x b^{i,j}_k(X^n_{t^n_k}, H_{k+1})Y_k\|_{V,p} \\
&\leqslant CM_{2p}(Z)^{1/p}\|\psi\|_{1,3,\infty}\frac{1}{n}\sum_{k=0}^{m-1}\|Y_k\|_{V,p},
\end{aligned}$$

and in the same way $\|\sum_{k=0}^{m-1}\delta^n_{k+1}\nabla_x\tilde{b}_k(X^n_{t^n_k}, H_{k+1}, \delta^n_{k+1})Y_k\|_{V,p} \leqslant C\|\psi\|_{1,3,\infty}\sum_{k=0}^{m-1}\|Y_k\|_{V,p}/n$. We gather all the terms and we obtain

$$\begin{aligned}
\|Y_m\|_{V,p} \leqslant\ & \|Y_0\|_{V,p} + CM_{2p}(Z)^{1/p}\|\psi\|_{1,3,\infty}(\frac{1}{n}\sum_{k=0}^{m-1}\|Y_k\|^2_{V,p})^{1/2} \\
& + C(M_p(Z)^{1/p}\sup_{k\leqslant n-1}\|\alpha^i_k\|_{V,p} + \frac{m_*^{1/p}}{r_*}\sup_{k\leqslant n-1}\|\beta^i_k\|_{V,p}) + \|\Gamma_m\|_{V,p}
\end{aligned}$$

Using Gronwall's lemma we obtain (1.106).
Step 2. Let

$$H = \{h : \{1,\ldots,n\}\times\{1,\ldots,N\} \to \mathbb{R} : |h|^2_H = \sum_{k=1}^n\sum_{i=1}^N h^2(k,i) < \infty\}.$$

so that $DX^n_{t^n_m} \in H^d$. We are going to prove that

$$\sup_{m\leqslant n}\|DX^n_{t^n_m}\|_{H^d,p} \leqslant CM_{2p}(Z)^{1/p}\|\psi\|_{1,3,\infty}\exp(CM_{2p}(Z)^{2/p}\|\psi\|^2_{1,3,\infty}). \tag{1.107}$$

For $h \in H$ we denote

$$D_hF = \langle DF, h\rangle = \sum_{k=1}^n\sum_{i=1}^N h(k,i)D_{(k,i)}F.$$

Since

$$D_{(r,j)}H_k^i = \frac{1}{\sqrt{n}}\delta_{r,k}\delta_{j,i}\chi_k,$$

we use (1.70) to obtain

$$\begin{aligned}
D_h X^n_{t^n_{k+1}} &= D_h X^n_{t^n_k} + A_k D_h X^n_{t^n_k} + \frac{1}{\sqrt{n}}\sum_{i=1}^N \chi_{k+1} h(k+1,i) a_k^i(X^n_{t^n_k}) \\
&+\frac{1}{\sqrt{n}}\sum_{i,j=1}^N \chi_{k+1}(h(k+1,i)H^j_{k+1} + h(k+1,j)H^i_{k+1}) b_k^{i,j}(X^n_{t^n_k}, H_{k+1}) \\
&+\frac{1}{\sqrt{n}}\sum_{i,j,q=1}^N \chi_{k+1} H^i_{k+1}H^j_{k+1}\partial_{z^q} b_k^{i,j}(X^n_{t^n_k}, H_{k+1}) h(k+1,q) \\
&+\frac{1}{\sqrt{n}}\chi_{k+1}\delta^n_{k+1}\sum_{q=1}^N \partial_{z^q}\tilde{b}_k(X^n_{t^n_k}, H_{k+1}, \delta^n_{k+1}) h(k+1,q)
\end{aligned}$$

Iterating this formula over k we obtain

$$D_h X^n_{t^n_m} = \sum_{k=0}^{m-1} A_k D_h X^n_{t^n_k} + \langle h, \Gamma_m\rangle$$

with $\Gamma_m(k,i) = 0$ for $k > m$ and, for $k \leqslant m$

$$\begin{aligned}
\Gamma_m(k,i) =& \frac{\chi_k}{\sqrt{n}}\big(a^i_{k-1}(X^n_{t^n_{k-1}}) + \sum_{j=1}^N H^j_k b^{i,j}_{k-1}(X^n_{t^n_{k-1}}, H_k) + \sum_{j,l=1}^N H^j_k H^l_k \partial_{z^i} b^{l,j}_{k-1}(X^n_{t^n_{k-1}}, H_k) \\
&+\delta^n_k \partial_{z^i}\tilde{b}_{k-1}(X^n_{t^n_{k-1}}, H_k, \delta^n_k)\big).
\end{aligned}$$

One has

$$|\Gamma_m|^2_{H^d} = \sum_{k=1}^m\sum_{i=1}^N |\Gamma_m(k,i)|^2 \leqslant C\|\psi\|^2_{1,3,\infty}\frac{1}{n}\sum_{k=1}^n (1+|Z_k|^4)$$

so, using (1.106) (with V replaced by H^d and $\alpha_k = \beta_k = 0$), we obtain

$$\begin{aligned}
\sup_{m\leqslant n}\|DX^n_{t^n_m}\|_{H^d,p} &\leqslant C\sup_{m\leqslant n}\|\Gamma_m\|_{H^d,p}\exp(CM_{2p}(Z)^{2/p}\|\psi\|^2_{1,3,\infty}) \\
&\leqslant CM_{2p}(Z)^{1/p}\|\psi\|_{1,3,\infty}\exp(CM_{2p}(Z)^{2/p}\|\psi\|^2_{1,3,\infty}).
\end{aligned}$$

Step 3. We estimate the derivatives of Y_m, solution of (1.103). We have

$$DY_m = \sum_{k=0}^{m-1} A_k DY_k + \sum_{i=1}^N\sum_{k=0}^{m-1} H^i_{k+1}\overline{\alpha}^i_k + \sum_{i=1}^N\sum_{k=0}^{m-1} LH^i_{k+1}\overline{\beta}^i_k + \overline{\Gamma}_m$$

with

$$\begin{aligned}
\overline{\alpha}^i_k &= \nabla_x\nabla_x a^i_k(X^n_{t^n_k}) DX^n_{t^n_k} Y_k + D\alpha^i_k, \\
\overline{\beta}^i_k &= D\beta^i_k
\end{aligned}$$

and

$$\begin{aligned}\overline{\Gamma}_m &= \sum_{k=0}^{m-1}\sum_{i=1}^{N}\nabla_x a_k^i(X_{t_k^n}^n)DH_{k+1}Y_k + \frac{1}{2}\sum_{i,j=1}^{N}\sum_{k=0}^{m-1} D(H_{k+1}^iH_{k+1}^j\nabla_x b_k^{i,j}(X_{t_k^n}^n,H_{k+1}))Y_k\\ &+ \sum_{k=0}^{m-1}\delta_{k+1}^n D(\nabla_x\tilde{b}_k(X_{t_k^n}^n,H_{k+1},\delta_{k+1}^n))Y_k + \sum_{i=1}^{N}\sum_{k=0}^{m-1}\alpha_k^i DH_{k+1}^i + \sum_{i=1}^{N}\sum_{k=0}^{m-1}\beta_k^i DLH_{k+1}^i + D\Gamma_m.\end{aligned}$$

Notice that DY_m is a process with values in H^d. We will prove that

$$\begin{aligned}&M_p(Z)^{1/p}C_{0,p}(\overline{\alpha},0,0) + \frac{m_*^{1/p}}{r_*}C_{0,p}(0,\overline{\beta},0) + C_{0,p}(0,0,\overline{\Gamma}) \leqslant \\ &\qquad CM_{4p}(Z)^{1/p}(M_{2p}(Z)^{1/2p}C_{0,2p}(\alpha,0,\Gamma) + \frac{m_*^{1/2p}}{r_*}(1+r_*^{-1})C_{0,2p}(0,\beta,0))\\ &\qquad \times\|\psi\|_{1,4,\infty}^2\exp(CM_{4p}(Z)^{1/p}\|\psi\|_{1,4,\infty}^2) + M_p(Z)^{1/p}C_{1,p}(\alpha,0,\Gamma) + \frac{m_*^{1/p}}{r_*}C_{1,p}(0,\beta,0).\end{aligned}\tag{1.108}$$

Once (1.108) is proved, the whole proof is concluded. Indeed, using (1.108) and the result from the first step (that is (1.105) with $q=0$ and Y_m replaced by DY_m), we obtain (1.105) with $q=1$. Consequently, using recursively the same reasoning we obtain (1.105) for every $q\in\mathbb{N}$. We estimate each of the terms which appear in the right hand side of (1.108). First, we write

$$\begin{aligned}\|\nabla_x\nabla_x a_k^i(X_{t_k^n}^n)DX_{t_k^n}^nY_k\|_{H^d,p} &\leqslant C\|\psi\|_{1,3,\infty}\big\||DX_{t_k^n}^n|_{H^d}|Y_k|_V\big\|_p \leqslant C\|\psi\|_{1,3,\infty}\|DX_{t_k^n}^n\|_{H^d,2p}\|Y_k\|_{V,2p}\\ &\leqslant CM_{4p}(Z)^{1/2p}\|\psi\|_{1,3,\infty}^2(M_{2p}(Z)^{1/2p}C_{0,2p}(\alpha,0,0)\\ &+ \frac{m_*^{1/2p}}{r_*}C_{0,2p}(0,\beta,0) + C_{0,2p}(0,0,\Gamma))\exp(CM_{4p}(Z)^{1/p}\|\psi\|_{1,3,\infty}^2),\end{aligned}$$

the last inequality being a consequence of (1.106) and (1.107). It follows that

$$\begin{aligned}\|\overline{\alpha}_k^i\|_{H^d,p} \leqslant\, &CM_{4p}(Z)^{1/2p}(M_{2p}(Z)^{1/2p}C_{0,2p}(\alpha,0,\Gamma) + \frac{m_*^{1/2p}}{r_*}C_{0,2p}(0,\beta,0))\\ &\times\|\psi\|_{1,3,\infty}^2\exp(CM_{4p}(Z)^{1/p}\|\psi\|_{1,3,\infty}^2) + C_{1,p}(\alpha,0,0).\end{aligned}$$

And

$$\|\overline{\beta}_k^i\|_{H^d,p} = \|D\beta_k^i\|_{H^d,p} \leqslant C_{1,p}(0,\beta,0).$$

We analyse now $\overline{\Gamma}_m$. We treat first $I_m := \sum_{k=0}^{m-1}\beta_k^i DLH_{k+1}^i$. Since $\beta_k^i D_{(p,j)}LH_{k+1}^i = 0$ if $p\neq k+1$, we obtain

$$|I_m|_{H^d}^2 \leqslant \sum_{j=1}^{N}\sum_{k=0}^{m-1}|D_{(k+1,j)}LH_{k+1}^i|^2|\beta_k^i|_V^2$$

so that, using (1.55), and the independency of LH_{k+1} and β_k, we have

$$\begin{aligned}\||I_m|_{H^d}\|_p &= \||I_m|_{H^d}^2\|_{p/2}^{1/2} \leqslant (\sum_{j=1}^{N}\sum_{k=0}^{m-1}\big\||D_{(k+1,j)}LH_{k+1}^i|^2|\beta_k^i|_V^2\big\|_{p/2})^{1/2}\\ &= (\sum_{j=1}^{N}\sum_{k=0}^{m-1}\big\||D_{(k+1,j)}LH_{k+1}^i||\beta_k^i|_V\big\|_p^2)^{1/2} = (\sum_{j=1}^{N}\sum_{k=0}^{m-1}\|D_{(k+1,j)}LH_{k+1}^i\|_p^2\big\||\beta_k^i|_V\big\|_p^2)^{1/2}\\ &\leqslant \frac{Cm_*^{1/p}}{r_*}(1+r_*^{-1})\sup_{k\leqslant m-1}\big\||\beta_k^i|_V\big\|_p = \frac{Cm_*^{1/p}}{r_*}(1+r_*^{-1})\sup_{k\leqslant m-1}\|\beta_k^i\|_{V,p}.\end{aligned}$$

Since DH_k^i has properties which are similar to the ones of DLH_k^i, the same reasoning as above gives

$$\|\sum_{k=0}^{m-1}\alpha_k^i DH_{k+1}^i\|_{H^d,p}\leqslant C\sup_{k\leqslant m-1}\|\alpha_k^i\|_{V,p}$$

and we have

$$\begin{aligned}|\sum_{k=0}^{m-1}\nabla_x a_k^i(X_{t_k^n}^n)Y_k DH_{k+1}^i|_{H^d}^2 &\leqslant \|\psi\|_{1,2,\infty}^2\sum_{k=0}^{m-1}\sum_{j=1}^{N}|Y_k|_V^2|D_{k+1,j}H_{k+1}^i|^2\\ &\leqslant \frac{C}{n}\|\psi\|_{1,2,\infty}^2\sum_{k=0}^{n-1}|Y_k|_V^2.\end{aligned}$$

Using (1.106) and the triangle inequality, we obtain

$$\begin{aligned}\|\sum_{k=0}^{m-1}\nabla_x a_k^i(X_{t_k^n}^n)Y_k DH_{k+1}^i\|_{H^d,p} &\leqslant C\|\psi\|_{1,2,\infty}(n^{-1}\sum_{k=0}^{n-1}\|Y_k\|_{V,p}^2)^{1/2}\\ &\leqslant C\big(M_p(Z)^{1/p}C_{0,p}(\alpha,0,\Gamma)+\frac{m_*^{1/p}}{r_*}C_{0,p}(0,\beta,0)\big)\\ &\quad\times\|\psi\|_{1,2,\infty}\exp(CM_{2p}(Z)^{2/p}\|\psi\|_{1,3,\infty}^2).\end{aligned}$$

We write now

$$\sum_{k=0}^{m-1}D(H_{k+1}^iH_{k+1}^j\nabla_x b_k^{i,j}(X_{t_k^n}^n,H_{k+1}))Y_k=I+J$$

with

$$\begin{aligned}I &= \sum_{k=0}^{m-1}(H_{k+1}^iDH_{k+1}^j+H_{k+1}^jDH_{k+1}^i)\nabla_x b_k^{i,j}(X_{t_k^n}^n,H_{k+1})Y_k,\\ J &= \sum_{k=0}^{m-1}H_{k+1}^iH_{k+1}^jD(\nabla_x b_k^{i,j}(X_{t_k^n}^n,H_{k+1}))Y_k.\end{aligned}$$

We have

$$|I|_{H^d}^2\leqslant C\|\psi\|_{1,3,\infty}^2n^{-1}\sum_{k=0}^{m-1}(|H_{k+1}^i|^2+|H_{k+1}^j|^2)|Y_k|_V^2,$$

and using the independence between Y_k and H_{k+1}, it follows that

$$\|I\|_{H^d,p}\leqslant Cn^{-1/2}\big(M_p(Z)^{1/p}C_{0,p}(\alpha,0,\Gamma)+\frac{m_*^{1/p}}{r_*}C_{0,p}(0,\beta,0)\big)M_p(Z)^{1/p}\|\psi\|_{1,3,\infty}\exp(CM_{2p}(Z)^{2/p}\|\psi\|_{1,3,\infty}^2).$$

Considering the estimates of $DX^n_{t^n_k}$, we obtain in a similar way

$$
\begin{aligned}
\|J\|_{H^d,p} &\leqslant Cn^{-1}(M_p(Z)^{1/p}C_{0,p}(\alpha,0,\Gamma)+\frac{m_*^{1/p}}{r_*}C_{0,p}(0,\beta,0))M_{2p}(Z)^{1/p}\|\psi\|_{1,3,\infty}\exp(CM_{2p}(Z)^{2/p}\|\psi\|^2_{1,3,\infty})\\
&+Cn^{-1/2}(M_{2p}(Z)^{1/2p}C_{0,2p}(\alpha,0,\Gamma)+\frac{m_*^{1/2p}}{r_*}C_{0,2p}(0,\beta,0))M_{4p}(Z)^{1/p}\\
&\times\|\psi\|^2_{1,4,\infty}\exp(CM_{4p}(Z)^{1/p}\|\psi\|^2_{1,3,\infty})\\
&\leqslant Cn^{-1/2}(M_{2p}(Z)^{1/2p}C_{0,2p}(\alpha,0,\Gamma)+\frac{m_*^{1/2p}}{r_*}C_{0,2p}(0,\beta,0))M_{4p}(Z)^{1/p}\\
&\times\|\psi\|^2_{1,4,\infty}\exp(CM_{4p}(Z)^{1/p}\|\psi\|^2_{1,3,\infty})
\end{aligned}
$$

It follows that a similar estimate holds for $\sum_{k=0}^{m-1}D(H^i_{k+1}H^j_{k+1}\nabla_x b_{i,j}(X^n_{t^n_k}))Y_k$ as for J. Finally, in the same way, we obtain

$$
\begin{aligned}
&\|\sum_{k=0}^{m-1}\delta^n_{k+1}D(\nabla_x\tilde{b}_k(X^n_{t^n_k},H_{k+1},\delta^n_{k+1}))Y_k\|_{H^d,p}\leqslant\\
&\qquad Cn^{-1/2}(M_{2p}(Z)^{1/2p}C_{0,2p}(\alpha,0,\Gamma)+\frac{m_*^{1/2p}}{r_*}C_{0,2p}(0,\beta,0))\|\psi\|^2_{1,4,\infty}\exp(CM_{4p}(Z)^{1/p}\|\psi\|^2_{1,3,\infty}).
\end{aligned}
$$

We gather all these terms and we obtain (1.108).

□

Now, we are in a position to prove Theorem 1.4.1. For the reader's convenience we recall the statement of this result.

Theorem 1.4.3. *For every $q,q'\in\mathbb{N}$, $q'\leqslant q$, and $p\geqslant 2$ there exists some constants $l\in\mathbb{N}^*$, $C\geqslant 1$ (depending on $r_*,\varepsilon_*,m_*,q,p$ and the moments of Z but not on n) such that*

$$
\sup_{t\in\pi^T_{T,n}}\sup_{0\leqslant|\alpha|\leqslant q-q'}\|\partial^\alpha_x X^n_t(x)\|_{q',p}\leqslant C\mathfrak{K}_{q+2}(\psi)^l, \tag{1.109}
$$

$$
\sup_{t\in\pi^T_{T,n}}\|LX^n_t\|_{q,p}\leqslant C\mathfrak{K}_{q+4}(\psi)^l. \tag{1.110}
$$

where $\mathfrak{K}_r(\psi)$ is defined in (1.69) and is given by

$$
\mathfrak{K}_r(\psi)=(1+\|\psi\|_{1,r,\infty})\exp(\|\psi\|^2_{1,3,\infty}).
$$

Proof. . We estimate first $\|X^n_t\|_{q,p}$. We have already checked that

$$
DX^n_{t^n_m}=\sum_{k=0}^{m-1}A_kDX^n_{t^n_k}+\Gamma_m
$$

with

$$
\begin{aligned}
\Gamma_m(k,i)=&\mathbb{1}_{\{k\leqslant m\}}\frac{\chi_k}{\sqrt{n}}(a^i_{k-1}(X^n_{t^n_{k-1}})+\sum_{j=1}^N H^j_k b^{i,j}_{k-1}(X^n_{t^n_{k-1}},H_k)+\sum_{j,l=1}^N H^j_kH^l_k\partial_{z^i}b^{l,j}_{k-1}(X^n_{t^n_{k-1}},H_k)\\
&+\delta^n_k\partial_{z^i}\tilde{b}_{k-1}(X^n_{t^n_{k-1}},H_k,\delta^n_k)).
\end{aligned}
$$

Using (1.105), the only thing to prove is that $\|\Gamma_m\|_{q-1,p} \leqslant C\mathfrak{K}_{q+2}(\psi)^l$. We have already done it for the first order derivatives (that is $q=1$). For higher order derivatives, the proof follows the same line (using a recurrence argument).
Now, we study $\nabla_x X^n_t(x)$ which solves the equation

$$\nabla_x X^n_{t^n_m}(x) = I + \sum_{k=1}^{m-1} A_k \nabla_x X^n_{t^n_k}(x).$$

This equation is similar to (1.103) so the upper bound of $\|\nabla_x X^n_{t^n_m}(x)\|_{q,p}$ follows from (1.106). For higher order derivatives the reasoning is the same.
Let us now deal with LX^n_t. Notice that $\left\langle DH^j_k, DH^i_k \right\rangle = 0$ for $i \neq j$. Then, using the computational rules (see (1.46)), we obtain

$$LX^n_{t^n_{k+1}} = A_k LX^n_{t^n_k} + \sum_{i=1}^N H^i_{k+1}\alpha^i_k + \sum_{i=1}^N LH^i_{k+1}\beta^i_k + \sum_{i,j=1}^N \gamma^{i,j}_k$$

with

$$\alpha^i_k = \sum_{l,r=1}^d \partial_{x_l}\partial_{x_r} a^i_k(X^n_{t^n_k})\langle (DX^n_{t^n_k})^r, (DX^n_{t^n_k})^l\rangle, \qquad \beta^i_k = a^i_k(X^n_{t^n_k})$$

and

$$\begin{aligned}
\gamma^{i,j}_k \;=\; & \frac{1}{2} LH^i_{k+1}H^j_{k+1} b^{i,j}_k(X^n_{t^n_k}, H_{k+1}) + \frac{1}{2} LH^i_{k+1}H^j_{k+1} b^{i,j}_k(X^n_{t^n_k}, H_{k+1}) \\
& + \frac{1}{2} H^i_{k+1}H^j_{k+1}\Big(\sum_{l,r=1}^d \partial_{x_l}\partial_{x_r} b^{i,j}_k(X^n_{t^n_k}, H_{k+1})\langle (DX^n_{t^n_k})^l, (DX^n_{t^n_k})^r\rangle + \sum_{r=1}^N \partial_{z_r} b^{i,j}_k(X^n_{t^n_k}, H_{k+1}) LH^r_{k+1} \\
& + \frac{\chi_{k+1}}{n}\sum_{r=1}^N \partial^2_{z_r} b^{i,j}_k(X^n_{t^n_k}, H_{k+1})\Big) + 1\!\!1_{i=j}\frac{\chi_{k+1}}{n} b^{i,i}_k(X^n_{t^n_k}, H_{k+1}) \\
& + \frac{\chi_{k+1}}{n}(H^i_{k+1}\partial_{z_j} b^{i,j}_k(X^n_{t^n_k}, H_{k+1}) + H^j_{k+1}\partial_{z_i} b^{i,j}_k(X^n_{t^n_k}, H_{k+1})) \\
& + \frac{1}{2}\delta^n_{k+1}\Big(\sum_{l,r=1}^d \partial_{x_l}\partial_{x_r}\tilde{b}_k(X^n_{t^n_k}, H_{k+1}, \delta^n_{k+1})\langle (DX^n_{t^n_k})^l, (DX^n_{t^n_k})^r\rangle + \sum_{r=1}^N \partial_{z_r}\tilde{b}_k(X^n_{t^n_k}, H_{k+1}, \delta^n_{k+1}) LH^r_{k+1} \\
& + \frac{\chi_{k+1}}{n}\sum_{r=1}^N \partial^2_{z_r}\tilde{b}_k(X^n_{t^n_k}, H_{k+1}, \delta^n_{k+1})\Big).
\end{aligned}$$

We have

$$\|\alpha^i_k\|_{q,p} \leqslant C\|\psi\|_{1,q+3,\infty}\|X^n_{t^n_k}\|^2_{q+1,p} \leqslant C\mathfrak{K}_{q+3}(\psi)^l$$

and a similar estimate holds for $\|\beta^i_k\|_{q,p}$. Moreover, we have $\Gamma_m = \sum_{i,j=1}^N \sum_{k=0}^{m-1} \gamma^{i,j}_k$ so we have to analyse each of the terms in $\gamma^{i,j}_k$. We look first at

$$\begin{aligned}
\|LH^i_{k+1}H^j_{k+1} b^{i,j}_k(X^n_{t^n_k}, H_{k+1})\|_{q,p} &\leqslant \|LH^i_{k+1}H^j_{k+1}\|_{q,2p}\|b^{i,j}_k(X^n_{t^n_k}, H_{k+1})\|_{q,2p} \\
&\leqslant \|LH^i_{k+1}\|_{q,4p}\|H^j_{k+1}\|_{q,4p}\|\psi\|^l_{1,q+2,\infty}(\|X^n_{t^n_k}\|^l_{q,2p} + \|H^j_k\|^l_{q,2p}) \\
&\leqslant C\mathfrak{K}_{q+2}(\psi)^l/n.
\end{aligned}$$

The other terms in $\gamma_k^{i,j}$ verify similar estimates. So we obtain

$$\|\Gamma_m\|_{q,p} \leqslant \sum_{i,j=1}^{N} \sum_{k=0}^{m-1} \|\gamma_k^{i,j}\|_{q,p} \leqslant C\mathfrak{K}_{q+4}(\psi)^l.$$

We conclude that

$$C_{q,p}(\alpha,\beta,\Gamma) \leqslant C\mathfrak{K}_{q+4}(\psi)^l$$

and the proof is competed.

□

1.5 The Ninomiya Victoir scheme

We illustrate Theorem 1.4.2 when X^n is the Ninomiya Victoir scheme for a diffusion process. This is a variant of the result already obtained by Kusuoka [44] in the case where Z_k has a Gaussian distribution (and so the standard Malliavin calculus is available). Since in our paper Z_k has an arbitrary distribution (except for the property (1.32)), our result may be seen as an invariance principle as well. We consider the d dimensional diffusion process

$$dX_t = \sum_{i=1}^{N} V_i(X_t) \circ dW_t^i + V_0(X_t)dt \tag{1.111}$$

with $V_0, V_i \in \mathcal{C}_b^\infty(\mathbb{R}^d;\mathbb{R}^d),\ i = 1,\ldots,N$ and $W = (W^1,\ldots,W^N)$ a standard Brownian motion and $\circ dW_t^i$ denotes the Stratonovich integral with respect to W^i. The infinitesimal operator of this Markov process is given by

$$A = V_0 + \frac{1}{2}\sum_{k=1}^{N} V_k^2 \tag{1.112}$$

with the notation $Vf(x) = \langle V(x), \nabla f(x)\rangle$. Let us define $\exp(V)(x) := \Phi_V(x,1)$ where Φ_V solves the deterministic equation

$$\Phi_V(x,t) = x + \textstyle\int_0^t V(\Phi_V(x,s))ds. \tag{1.113}$$

By a change of variables, it is possible to show that $\Phi_{\varepsilon V}(x,t) = \Phi_V(x,\varepsilon t)$, so we have

$$\exp(\varepsilon V)(x) := \Phi_{\varepsilon V}(x,1) = \Phi_V(x,\varepsilon).$$

We also notice that the semigroup of the above Markov process is given by $P_t^V f(x) = f(\Phi_V(x,t))$ and has the infinitesimal operator $A_V f(x) = Vf(x)$. In particular the relation $P_t^V A_V = A_V P_t^V$ reads

$$Vf(\Phi_V(x,t)) = A_V P_t^V f = P_t^V A_V f = \langle V(x), \nabla_x\left(f(\Phi_V(x,t)\right)\rangle.$$

Using m times Dynkin's formula $P_t^V f(x) = f(x) + \int_0^t P_s^V A_V f(x)ds$ we obtain

$$f(\Phi_V(x,t))) = f(x) + \sum_{r=1}^{m} \frac{t^r}{r!} V^r f(x) + \frac{1}{m!}\int_0^t (t-s)^m V^{m+1} P_s^V f(x)ds. \tag{1.114}$$

We present now the Ninomiya Victoir scheme. We consider a sequence $\rho_k,\ k \in \mathbb{N}$ of independent Bernoulli random variables and we define $\psi_k : \mathbb{R}^d \times \mathbb{R}^{N+1} \to \mathbb{R}^d$ in the following way

$$\psi_k(x, w^1, w^0) = \exp(w^0 V_0) \circ \exp(w^{1,1} V_1) \circ \cdot \circ \exp(w^{1,N} V_N) \circ \exp(w^0 V_0)(x), \text{ if } \rho_k = 1, \quad (1.115)$$
$$\psi_k(x, w^1, w^0) = \exp(w^0 V_0) \circ \exp(w^{1,N} V_N) \circ \cdot \circ \exp(w^{1,1} V_1) \circ \exp(w^0 V_0)(x), \text{ if } \rho_k = -1. \quad (1.116)$$

The Ninomiya Victoir scheme uses these functions with $w_k^0 = T/2n$ and $w_k^{1,i} = \sqrt{T} Z_k^i/\sqrt{n}$, for $i = 1, \ldots, N$. Moreover $Z_k^i,\ i = 1, \ldots, d,\ k \in \mathbb{N}^*$ are independent random variables which verify (1.32) and moreover satisfy the following moment conditions:

$$\mathbb{E}[Z_k^i] = \mathbb{E}[(Z_k^i)^3] = \mathbb{E}[(Z_k^i)^5] = 0, \qquad \mathbb{E}[(Z_k^i)^2] = 1, \qquad \mathbb{E}[(Z_k^i)^4] = 6. \quad (1.117)$$

In the original paper of Ninomiya Victoir, the random variables Z_k^i are standard normally distributed, and then verify (1.32). The new point here is that we do not require that Z_k follows this particular law anymore but only the weaker assumptions (1.32) and (1.117). We recall that $t_k^n = Tk/n$. One step of our scheme is given by

$$X^n_{t^n_{k+1}} = \psi_k(X^n_{t^n_k}, w^1_{k+1}, w^0_{k+1}). \quad (1.118)$$

We have the first following result.

Theorem 1.5.1. *There exists some universal constants $l \in \mathbb{N}^*$, $C \geqslant 1$ such that for every $f \in \mathcal{C}_b^6(\mathbb{R}^d)$, we have*

$$\sup_{t \in \pi^T_{T,n}} |\mathbb{E}[f(X_t)] - \mathbb{E}[f(X^n_t)]| \leqslant C C_6(V)^l \|f\|_{6,\infty}/n^2 \quad (1.119)$$

with $C_q(V) := \sup_{i=0,.,N} \|V_i\|_{q,\infty}$.

Remark 1.5.1. *The same estimate has already been proved by Alfonsi [3] using short time expansions on the solution of the Feynman Kac partial dfferential equation associated to the diffusion process.*

Under an ellipticity condition we are able to give an estimate of the total variation distance between a diffusion process of the form (1.111) and its Ninomiya Victoir scheme.

Theorem 1.5.2. *We assume that*

$$\inf_{|\xi|=1} \sum_{i=1}^N \langle V_i(x), \xi \rangle^2 \geqslant \lambda_* > 0 \qquad \forall x \in \mathbb{R}^d. \quad (1.120)$$

Let $S \in (0, T/2)$. Then there exists $n_0 \in \mathbb{N}^*$ such that for every $n \geqslant n_0$, there exists $l \in \mathbb{N}^*$, $C \geqslant 1$ such that for every bounded and measurable function $f : \mathbb{R}^d \to \mathbb{R}$,

$$\sup_{t \in \pi^{2S,T}_{T,n}} |\mathbb{E}[f(X_t)] - \mathbb{E}[f(X^n_t)]| \leqslant C \frac{C_6(V)^l \mathfrak{K}_9(\psi)^l}{(\lambda_* S)^{42}} \|f\|_\infty /n^2. \quad (1.121)$$

Remark 1.5.2. *This estimate has already been proved by Kusuoka [44] (with a different approach). He considers a much more general non degeneracy assumptions (of Hörmander type) and uses Malliavin calculus in order to prove his result. Here the noise Z_k^i is no more Gaussian so the standard Malliavin calculus does not work anymore, but, since we have the property (1.32), we may use the abstract integration by parts formula introduced in Section 1.3.*

Proof of Theorem 1.5.1. We have to show $E_n(3,6)$ (see (1.16)) and (1.15) for Q^n. Indeed, the proof will then follow from Proposition 1.2.1. First, we notice that (1.15) is satisfied with $q=6$ for the semigroup Q^n using Theorem 1.4.1 (see (1.71)). Now, we focus on the proof of $E_n(3,6)$. In order to simplify the notations, we fix $T=1$ without loss of generality. We denote

$$\mathcal{T}_0 f(x)=\mathcal{T}_{N+1}f(x)=f(\exp(\frac{1}{2n}V_0)(x)),\quad \mathcal{T}_i f(x)=f(\exp(\frac{Z}{\sqrt{n}}V_1)(x)), i=1,\ldots,N.$$

Notice that, with the notation introduced in the beginning of this section, $\mathcal{T}_i f(x)=P_i^{U_i}f(x)$ with $U_i=ZV_i/\sqrt{n}$, if $i=1,\ldots,N$ and $U_0=U_{N+1}=V_0/(2n)$. Using (1.114) with $t=1$ and $V=U_i, i=1,\ldots,N$ we obtain

$$\mathcal{T}_i f(x)=f(x)+\sum_{r=1}^{m}\frac{Z^r}{n^{r/2}}\frac{1}{r!}V_i^r f(x)+\frac{Z^{m+1}}{n^{(m+1)/2}}R_{m+1,i}f(x) \tag{1.122}$$

with

$$R_{m+1,i}f(x)=\frac{1}{m!}\int_0^1(1-\lambda)^m V_i^{m+1}P_\lambda^{U_i}f(x)d\lambda \tag{1.123}$$

and we recall that $P_\lambda^{U_i}f(x)=f(\exp(\lambda ZV_i/\sqrt{n}))$. We have a similar expansion if we put $V=V_0/(2n)$ in (1.114). We aim to give an expansion of order 3 (with respect to $1/n$) for $\mathbb{E}[f(\psi_k(x,w_{k+1}^1,w_{k+1}^0)]$ (see (1.124) below). In order to do it, we replace each $\mathcal{T}_i, i=1,\ldots,N$, with an expansion of order $m\leqslant 5$ given above with $Z=Z_{k+1}^i$ (and we proceed in the same way when $V=V_0/(2n)$). Then, we calculate the products of the miscellaneous expansions, each with a well chosen order m such that there is no term with factor n^{-r}, $r>3$, appearing in those products. Moreover, all the terms containing n^{-3} go in the remainder. The last step consists in computing the expectancy. We notice that $\mathbb{E}[P_t^{U_i}]=P_t^{V_i^2/(2n)}$ and $\mathbb{E}[(Z_{k+1}^i)^r]=0$ for odd $r\leqslant 5$. Finally, since $\mathbb{E}[(Z_{k+1}^i)^2]=1$, $\mathbb{E}[(Z_{k+1})^4]=6$, the calculus is completed and we obtain:

$$\begin{aligned}\mathbb{E}[f(\psi_k(x,w_{k+1}^1,w_{k+1}^0)] &= \mathbb{E}[\mathcal{T}_0\mathcal{T}_1\ldots\mathcal{T}_{N+1}f(x)] \qquad (1.124)\\ &= f(x)+\frac{1}{n}(V_0f(x)+\frac{1}{2}\sum_{i=1}^{N}V_i^2f(x))+\frac{1}{2n^2}V_0^2f(x)+\frac{1}{8n^2}\sum_{i=1}^{N}V_i^4f(x)\\ &\quad+\frac{1}{4n^2}\sum_{i<j}V_i^2V_j^2f(x)+\frac{1}{4n^2}\sum_{i=1}^{N}(V_0V_i^2f(x)+V_i^2V_0f(x))+\frac{1}{n^3}Rf(x).\end{aligned}$$

The remainder R is a sum of terms of the following form:

$$C\mathcal{T}_{0,\alpha_0},\ldots,\mathcal{T}_{N+1,\alpha_{N+1}}f(x) \tag{1.125}$$

with $\alpha=(\alpha_0,\ldots,\alpha_{N+1})\in\{0,\ldots,3\}^{N+2}$, $|\alpha|=\alpha_0+\ldots+\alpha_{N+1}=3$, and using the notation given in (1.123),

$$\begin{aligned}&\mathcal{T}_{0,k},\mathcal{T}_{N+1,k}\in\{V_0^k,R_{k,0}\}, &&\mathcal{T}_{i,k}\in\{V_i^{2k},R_{2k,i}\}, i\in\{1,\ldots,N\}\quad k=0,\ldots,2,\\ &\mathcal{T}_{0,3}=\mathcal{T}_{N+1,3}=R_{3,0}, &&\mathcal{T}_{i,3}=\mathbf{R}_{6,i}, i\in\{1,\ldots,N\},\end{aligned}$$

with for $i=1,\ldots,N$,

$$\mathbf{R}_{6,i}=\mathbb{E}[(Z^i)^6R_{6,i}]=\int_0^1(1-\lambda)^5\mathbb{E}[Z^6V_i^6P_\lambda^{U_1}f(x)]d\lambda.$$

It is easy to check that for every $g\in\mathcal{C}^{k+p}(\mathbb{R})$, we have the following property

$$\|\mathcal{T}_{i,k}g\|_{p,\infty}\leqslant CC_{2k+p}(V)^l\|g\|_{k+p,\infty}$$

for some constants $l \in \mathbb{N}^*$, $C \geqslant 1$. So, it follows that

$$\|Rf\|_\infty \leqslant CC_6(V)^l \|f\|_{6,\infty}. \tag{1.126}$$

We turn now to the diffusion process X_t. For any $t > 0$, we have the expansion

$$\mathbb{E}[f(X_t(x))] = P_t^A f(x) = f(x) + tAf(x) + \frac{t^2}{2}A^2 f(x) + \frac{t^3}{3!}R'_t f(x).$$

with

$$R'_t f(x) = t^{-1}\int_0^t P_\lambda^A A^3 f(x)(1-\lambda/t)^2 d\lambda. \tag{1.127}$$

We take $t = n^{-1}$ and make the difference between (1.127) and (1.124). All the terms cancel except for the remainders so we obtain

$$\begin{aligned}&\forall k \in \{0,\dots,n-1\},\\ &\qquad \mathbb{E}[f(X_{t^n_{k+1}})] - \mathbb{E}[f(X^n_{t^n_{k+1}}) \mid X_{t^n_k} = X^n_{t^n_k} = x] = (R'_{1/n}f(x)/3! - Rf(x))/n^3.\end{aligned} \tag{1.128}$$

We clearly have $\|R'_{1/n}f\|_\infty \leqslant CC_6(V)^l\|f\|_{6,\infty}$. This, together with (1.126) completes the proof. □

Proof of Theorem 1.5.2. This will be a consequence of Theorem 1.4.2 as soon as we check that the ellipticity assumption (1.82) holds true. We fix k and we look at $\psi_k(x, w^1, w^0)$ defined in (1.116). We suppose that $\rho_k = 1$ (the proof for $\rho_k = -1$ is similar). We denote $w^1 = (w^{1,1}, \cdots, w^{1,N}\}$ and $\tilde{w} = (w^1, w^0)$ with $w^0 \in \mathbb{R}_+$ and $T_i = i$ and we consider the process $x_t(w), 0 \leqslant t \leqslant K_{N+2}$ solution of the following equation:

$$\begin{aligned}
x_t(\tilde{w}) &= x + \frac{w_0}{2}\int_{T_0}^t V_0(x_s(\tilde{w}))ds, \qquad T_0 \leqslant t \leqslant T_1,\\
x_t(\tilde{w}) &= x_{T_i}(\tilde{w}) + w^{1,i}\int_{T_k}^t V_i(x_s(\tilde{w}))ds, \qquad T_i \leqslant t \leqslant T_{i+1}, \qquad i = 1,\dots,N,\\
x_t(\tilde{w}) &= x_{T_{N+1}}(\tilde{w}) + \frac{w^0}{2}\int_{T_{N+1}}^t V_0(x_s(\tilde{w}))ds, \qquad T_{N+1} \leqslant t \leqslant T_{N+2}.
\end{aligned}$$

Then, $\psi_k(x,\tilde{w}) = x_{T_{N+2}}(\tilde{w})$ and consequently for $r \in \{1,\dots,N\}$, we have $\partial_{w^{1,r}}\psi_k(x,\tilde{w}) = \partial_{w^{1,r}}x_{T_{N+2}}(\tilde{w})$. Moreover $\partial_{w^{1,r}}x_t(\tilde{w}) = 0$ for $t \leqslant T_r$ and

$$\begin{aligned}
\partial_{w^{1,r}}x_t(\tilde{w}) = \partial_{w^{1,r}}x_{T_{r+1}}(\tilde{w}) + \sum_{i=r+1}^{N+1} w^{1,i}\int_{T_i\vee t}^{T_{i+1}\vee t} \nabla V_i(x_s(\tilde{w}))\partial_{w^{1,r}}x_s(\tilde{w})ds\\
+\frac{w_0}{2}\int_{T_{N+1}\vee t}^t \nabla V_0(x_s(\tilde{w}))\partial_{w^{1,r}}x_s(\tilde{w})ds,
\end{aligned}$$

for $t \geqslant T_{r+1}$, in particular for $t = T_{N+1}$. For $T_r < t \leqslant T_{r+1}$, $\partial_{w^{1,r}}x_t(\tilde{w})$ solves the equation

$$\partial_{w^{1,r}}x_t(\tilde{w}) = \int_{T_r}^t V_r(x_s(\tilde{w}))ds + w^{1,r}\int_{T_r}^t \nabla V_r(x_s(\tilde{w}))\partial_{w^{1,r}}x_s(\tilde{w})ds.$$

It follows that

$$\partial_{w^{1,r}}x_t(\tilde{w}) \mid_{\tilde{w}=0} = \int_{T_r}^t V_r(x_s(0))ds = V_r(x)(t - T_r).$$

Notice that $T_{r+1} - T_r = 1$. Then, we have

$$\partial_{w^{1,r}} x_{T_{N+2}}(\tilde{w}) \mid_{\tilde{w}=0} = \partial_{w^{1,r}} x_{T_{r+1}}(\tilde{w}) \mid_{\tilde{w}=0} = V_r(x).$$

and then, by (1.120),

$$\sum_{r=1}^{N} \langle \partial_{w^{1,r}} x_{T_{N+2}}(0), \xi \rangle^2 \geqslant \lambda_* |\xi|^2.$$

$\square$

1.6 Complement: Malliavin calculus with simple functionals approach

In Malliavin calculus, we assume that random variables $Z_k, k \in \mathbb{N}^*$ follow a Gaussian law. In this paper, we start with a more generic assumption. We assume the random variables $Z_k, k \in \mathbb{N}^*$ to follow an arbitrary law. As a consequence, we need to develop a dedicated differential calculus. The method we follow is however still inspired from standard Malliavin calculus since it is quite similar to the simple functionals approach. Basically, it consists in building a finite dimensional differential calculus for functionals of the increments of the Gaussian random variable. From this construction, we obtain the infinite dimensional differential formulas by making the time step of those increments tend to zero. It is worth noticing that the abstract Malliavin calculus which is developed in this paper is only finite dimensional. Nevertheless, it remains pretty close to the simple functional approach for finite dimensional case. That is why, in this section, we propose a brief presentation of standard Malliavin calculus from this perspective in both finite and infinite dimensional cases.

The notations we will use in this section prevail for this section only.

1.6.1 Finite dimensional differential calculus on the Wiener space

This section gives the main results of the simple functionals approach for finite dimensional case which inspired our work.

Simple functionals Let $n \in \mathbb{N}^*$ and $(W_t)_{t\geqslant 0}$ a standard Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{F}_t = \sigma(W_s, s \leqslant t)$. For $k \in \{0, 1, \dots, 2^n\}$, we denote $t_k^n = \frac{k}{2^n}$, $\Delta_k^{n,W} = W_{t_k^n} - W_{t_{k-1}^n}$ and $\Delta^{n,W} = \left(\Delta_1^{n,W}, .., \Delta_{2^n}^{n,W}\right)$. We define $\mathcal{S}_n$, the space of simple functionals of order n in the following way:

$$\mathcal{S}_n = \left\{F = f\left(\Delta_1^{n,W}, \dots, \Delta_{2^n}^{n,W}\right) : f \in \mathcal{C}_{pol}^{\infty}\left(\mathbb{R}^{2^n}, \mathbb{R}\right)\right\} \tag{1.129}$$

with for all $q \in \mathbb{N}$,

$$\mathcal{C}_{pol}^{q}\left(\mathbb{R}^{2n}, \mathbb{R}\right) = \left\{f \in \mathcal{C}^q, \forall i = 0, .., q, \alpha_i \in \{1, .., 2^n\}^i, \exists C > 0, e \in \mathbb{N}, \partial_{\alpha_i} f(x) \leq C(1 + |x|^e)\right\}$$

A first standard observation is that $\mathcal{S}_n \subset \mathcal{S}_{n+1}$. Then, we define the set of simple functionals: $\mathcal{S} = \cup_{n=1}^{\infty} \mathcal{S}_n$. Now, we are in a position to define the finite dimensional differential calculus for these simple functionals.

Malliavin operators Now we have defined the set of random variables on which we can apply the Malliaivin calculus, the next step is to introduce the operators that will appear in the integration by parts formulas. In the following, we define the Malliavin derivatives, the Malliavin covariance matrix and the the Ornstein Uhlenbeck operator. Before that, we introduce some notations. For $p \in \mathbb{N}^*$, we define $\mathrm{L}^p(\Omega; \mathrm{L}^2([0,1])) = \{(U_s)_{s\in[0,1]}, \mathbb{E}[\|U_s\|_{\mathrm{L}^2([0,1])}^p] < \infty\}$.

Definition 1.6.1. *We use the notations introduced in this section.*

A. *We define the Malliavin derivative operator $D : \mathcal{S} \to \mathrm{L}^2(\Omega; \mathrm{L}^2([0,1]))$. Let $F \in \mathcal{S}$. Then $F \in \mathcal{S}_n$ for a given $n \in \mathbb{N}$, and we define*

$$D_s F = \frac{\partial F}{\partial \Delta_s^{n,W}} \tag{1.130}$$

with $\Delta_s^{n,W} = \Delta_k^{n,W}$ for $s \in [t_{k-1}^n, t_k^n)$.

B. *Let $d \in \mathbb{N}^*$ and $F \in \mathcal{S}^d$. The Malliavin covariance matrix of F is defined by*

$$\sigma_F^{i,j} = \langle DF^i, DF^j \rangle = \int_0^1 D_s F^i D_s F^j ds. \tag{1.131}$$

C. *Finally, we introduce the Ornstein Ulhenbeck operator of $F = f(\Delta^{n,W})$*

$$LF = \sum_{k=1}^{2^n} \partial_k^2 f(\Delta^{n,W}) 2^{-n} - \partial_k f(\Delta^{n,W}) \Delta_k^{n,W}. \tag{1.132}$$

Duality and Integration by parts formulas In this framework we can obtain the following results

Proposition 1.6.1. *Duality formula*
Let $F, G \in \mathcal{S}$, then we have

$$\mathbb{E}[FLG] = \mathbb{E}[\langle DF, DG \rangle] = \mathbb{E}[GLF]. \tag{1.133}$$

Theorem 1.6.1. *Integration by parts formula*
Let $F \in \mathcal{S}^d$ and $G \in \mathcal{S}$ be such that $\mathbb{E}[(\det \sigma_F)^{-p}] < \infty$ for every $p \geqslant 1$. We denote $\gamma_F = \sigma_F^{-1}$. Then for every $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ and every $i = 1, \ldots, d$,

$$\mathbb{E}[\partial_i \phi(F) G] = \mathbb{E}[\phi(F) H_i(F, G)] \tag{1.134}$$

with

$$-H(F, G) = G\gamma_F LF + \langle D(G\gamma_F), DF \rangle \tag{1.135}$$

and

$$H_i(F, G) = -\sum_{j=1}^{d} G\gamma_F^{i,j} LF^j + \langle D(G\gamma_F^{i,j}), DF^j \rangle.$$

Moreover, for every multi index $\alpha = (\alpha_1, \ldots, \alpha_m) \in \{1, \ldots, d\}^m$

$$\mathbb{E}[\partial_\alpha \phi(F) G] = \mathbb{E}[\phi(F) H_\alpha(F, G)] \tag{1.136}$$

with $H_\alpha(F, G)$ defined by the recurrence relation $H_{(\alpha_1, \cdots, \alpha_m)}(F, G) = H_{\alpha_m}(F, H_{(\alpha_1, \ldots, \alpha_{m-1})}(F, G))$.

Both the aforementioned proposition and theorem are classical results in standard Malliavin calculus so we do not provide their proofs. However, it is important to notice that the proof of Theorem 1.6.1 relies on the study of the Gaussian random variable.

With these formulas, we have defined a finite dimensional differential calculus for functionals of the increments of the Brownian motion. This calculus is similar to the one which is required in our paper. Indeed, we study some Markov chains with form (1.1), where the random variables $Z_k/\sqrt{n}$, $k \in \mathbb{N}^*$, can be seen as independent increments of a random process. The main difference in our approach, is that we do not assume that Z_k has a Gaussian distribution. However, since the distribution of Z_k involves a smooth part, we use a similar approach and also obtain integration by parts formulas. Using those formulas, we obtain regularity properties of the semigroups and then the total variation convergence results.

In our study, the number of increments is fixed and then we obtain regularity properties of the semigroups built with this finite number of random variables Z_k. Even if the number of increments will be large in concrete applications (since we study convergence results), there is no necessary asymptotic procedure in our abstract Malliavin calculus, so we always work in finite dimensional case. However, the standard Malliavin calculus goes beyond this kind of framework. Indeed, the final purpose is to develop a differential calculus that relies on the path of the Brownian motion $(W_t)_{t\geqslant 0}$. In order to do it, it is necessary to develop an infinite dimensional calculus. Then, it will be possible to obtain integration by parts formulas for functionals of the Brownian path. We discuss this subject in the following.

1.6.2 Infinite dimensional differential calculus on the Wiener space

In this section, we present the integration by parts formulas that rely on the path of the Brownian motion.

Proposition 1.6.2. *We have the following properties:*

- ***A.*** *The space of simple functionals $\mathcal{S}$ is dense in $\mathrm{L}^2(\Omega)$*
- ***B.*** *The operator D is closable, that is: $\forall (F_i)_{i\in\mathbb{N}} \in \mathcal{S}$, if*

$$F_i \to 0 \quad \textit{in } \mathrm{L}^2(\Omega), \tag{1.137}$$

$$DF_i \to G \quad \textit{in } \mathrm{L}^2(\Omega;\mathrm{L}^2([0,1])), \tag{1.138}$$

then $G = 0$.

Using this result, we can define DF for all $F \in \mathrm{L}^2(\Omega)$ such that $\lim DF_i$ exists and then $DF = \lim DF_i$. We denote by $D^{1,2}$ this set of processes. This set is comparable to the Sobolev spaces that appear in distribution theory. As in this theory, we can recursively build higher order spaces $D^{k,p} \subset \mathrm{L}^p(\Omega;\mathrm{L}^2([0,1]^k))$ that involve the processes with Malliavin derivatives of order k. We denote

$$D^\infty = \cap_{k=1}^\infty \cap_{p=1}^\infty D^{k,p} \tag{1.139}$$

The set D^∞ is obviously larger than $\mathcal{S}$, but it is still possible to obtain integration by parts formulas in this infinite dimensional case.

Theorem 1.6.2. *Integration by parts formula*
Let $F \in (D^\infty)^d$ and $G \in D^\infty$ be such that $\mathbb{E}[(\det \sigma_F)^{-p}] < \infty$ for every $p \geqslant 1$. We denote $\gamma_F = \sigma_F^{-1}$. Then for every $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ and every $i = 1,\dots,d$

$$\mathbb{E}[\partial_i\phi(F)G] = \mathbb{E}[\phi(F)H_i(F,G)] \tag{1.140}$$

with

$$-H(F,G) = G\gamma_F LF + \langle D(G\gamma_F), DF\rangle \tag{1.141}$$

and

$$H_i(F,G) = -\sum_{j=1}^d G\gamma_F^{i,j}LF^j + \langle D(G\gamma_F^{i,j}), DF^j\rangle.$$

Moreover, for every multi index $\alpha = (\alpha_1,\dots,\alpha_m) \in \{1,\dots,d\}^m$

$$\mathbb{E}[\partial_\alpha\phi(F)G] = \mathbb{E}[\phi(F)H_\alpha(F,G)] \tag{1.142}$$

with $H_\alpha(F,G)$ defined by the recurrence relation $H_{(\alpha_1,\cdots,\alpha_m)}(F,G) = H_{\alpha_m}(F, H_{(\alpha_1,\dots,\alpha_{m-1})}(F,G))$.

1.6.3 Other perspectives

In this section, we have shown that we can obtain integration by parts formulas for processes that belong to $D^\infty \subset \mathrm{L}^2(\Omega)$ using the simple functional approach. However, this problem can be treated from other perspectives. Indeed, the basic objects can be different from simple functionals.

Another approach consists in using the Wiener chaos decomposition. We will not detail this method but we try to provide a quick overview. It relies on the following decomposition

$$\mathrm{L}^2(\Omega) = \bigoplus_{k=0}^{\infty} \mathcal{H}_k \tag{1.143}$$

with $\mathcal{H}_k$ the Wiener chaos of order k. In this way, any square integrable random variable can be represented as a sum of iterated stochastic integrals. The next step consists in defining the Malliavin derivatives of such iterated stochastic integral which are, in this approach, the basic objects. Finally, we use closability properties adapted to this framework and we obtain integration by parts formulas.

Chapter 2

Total variation convergence of order three of approximation scheme for one dimensional SDEs

Ce Chapitre fait l'objet d'un article qui sera soumis prochainement.

Abstract

In this paper, we study a third weak order scheme for diffusion processes which has been introduced by Alfonsi [3]. This scheme is built using cubature methods and is well defined under an abstract commutativity condition on the coefficients of the underlying diffusion process. Moreover, it has been proved in [3], that the third weak order convergence takes place for smooth test functions. First, we provide a necessary and sufficient explicit condition for the scheme to be well defined when we consider the one dimensional case. In a second step, we use a result from [10] and prove that, under an ellipticity condition, this convergence also takes place for the total variation distance with order 3. We also give an estimate of the density function of the diffusion process and its derivatives.

2.1 Introduction

In this paper, we study the total variation distance between a one dimensional diffusion process and a third weak order scheme based on a cubature method and introduced by Alfonsi [3]. In his work, Alfonsi proved that it converges with weak order three for smooth test functions with polynomial growth. We will show that the convergence also takes place with order three if we consider measurable and bounded test functions. In this case, we say that the total variation distance between the diffusion process and the scheme converges towards zero with order three. In order to do it, we will use a result from [10] based on en abstract Malliavin calculus introduced by Bally and Clément [9]. A main interest of this approach is that the random variables used to build the scheme are not necessarily Gaussian but belong to a class of random variables with no specific law. Consequently our result can be seen as an invariance principle.
Let us be more specific. We consider the $\mathbb{R}$-valued one dimensional Markov diffusion process

$$dX_t = V_0(X_t)dt + V_1(X_t) \circ dW_t, \tag{2.1}$$

with $V_i : \mathcal{C}_b^\infty(\mathbb{R}, \mathbb{R})$, $i = 0, 1$, $(W_t)_{t\geqslant 0}$ a one dimensional standard Brownian motion and $\circ dW_t$ the Stratonovich integral with respect to W_t. In this paper, we will study an approximation scheme for (2.1) which is defined on an homogeneous time grid. It is relevant to notice that the results we will obtain remain true for non homogeneous time grids, but we do not treat that case for sake of clarity. We fix $T>0$ and we denote $n\in\mathbb{N}^*$, the number of time step between 0 and T. Then, for $k\in\mathbb{N}$ we define $t^n_k=kT/n$ and we introduce the homogeneous time grid $\pi_{T,n}=\{t^n_k=kT/n,k\in\mathbb{N}\}$ and its bounded version $\pi^{\tilde{T}}_{T,n}=\{t\in\pi_{T,n},t\leqslant\tilde{T}\}$ for $\tilde{T}\geqslant 0$. Finally, for $S\in[0,\tilde{T})$ we will denote $\pi^{S,\tilde{T}}_{T,n}=\{t\in\pi^{\tilde{T}}_{T,n},t>S\}$. Now, for $t^n_k=kT/n$, we introduce the abstract $\mathbb{R}$-valued Markov chain

$$X^n_{t^n_{k+1}}=\psi_k(X^n_{t^n_k},\frac{Z_{k+1}}{\sqrt{n}},\delta^n_{k+1}),\qquad k\in\mathbb{N}, \tag{2.2}$$

where $\psi_k:\mathbb{R}\times\mathbb{R}\times\mathbb{R}_+\to\mathbb{R}$ is a smooth function such that $\psi_k(x,0,0)=x$, $Z_{k+1}\in\mathbb{R}^N$, $k\in\mathbb{N}$ is a sequence of independent and centered random variables and $\sup_{k\in\mathbb{N}^*}\delta^n_k\leqslant C/n$.

Before estimating the distance between X and X^n, we introduce some notations. For $f\in\mathcal{C}^\infty(\mathbb{R}^d)$ and for a multi-index $\alpha=(\alpha_1,\cdots,\alpha_d)\in\mathbb{N}^d$ we denote $|\alpha|=\alpha_1+...+\alpha_d$ and $\partial_\alpha f=\partial^\alpha_x f=\partial^{\alpha_1}_{x_1}...\partial^{\alpha_d}_{x_d}f(x)$. We include the multi-index $\alpha=(0,...,0)$ and in this case $\partial_\alpha f=f$. We will use the norms

$$\|f\|_{q,\infty}=\sup_{x\in\mathbb{R}^d}\sum_{0\leqslant|\alpha|\leqslant q}|\partial_\alpha f(x)|,\quad q\in\mathbb{N}. \tag{2.3}$$

In particular $\|f\|_{0,\infty}=\|f\|_\infty$ is the usual supremum norm and we will denote $\mathcal{C}^q_b(\mathbb{R}^d)=\{f\in\mathcal{C}^q(\mathbb{R}^d),\|f\|_{q,\infty}<\infty\}$.

A first standard result is the following: Let us assume that there exists $h>0$, $q\in\mathbb{N}$ such that for every test function $f\in\mathcal{C}^q_b(\mathbb{R})$, $k\in\mathbb{N}$ and $x\in\mathbb{R}$,

$$|\mathbb{E}[f(X_{t^n_{k+1}})-f(X^n_{t^n_{k+1}})|X_{t^n_k}=X^n_{t^n_k}=x]|\leqslant C\|f\|_{q,\infty}/n^{h+1}. \tag{2.4}$$

Then, we have

$$\sup_{t\in\pi^T_{T,n}}|\mathbb{E}[f(X_t)-f(X^n_t)]|\leqslant C\|f\|_{q,\infty}/n^h. \tag{2.5}$$

It means that $(X^n_{t_k})_{k\in\mathbb{N}}$ is an approximation scheme of weak order h for the Markov process $(X_t)_{t\geqslant 0}$ for the test functions $f\in\mathcal{C}^q_b(\mathbb{R};\mathbb{R})$. The value h thus measures the efficiency of the scheme whereas q stands for the required regularity on the test functions in order to obtain convergence with order h. This subject has already been widely studied in the literature and we point out some famous examples. However, the reader may notice that in all those works, the required order of regularity q is greater than one. Concerning the Euler scheme for diffusion processes, the result (2.5), with $h=1$, has initially been proved in the seminal papers of Milstein [55] and of Talay and Tubaro [68] (see also [39]). Since then, various situations have been studied: Diffusion processes with jumps (see [63], [36]) or diffusion processes with boundary conditions (see [30], [18], [31]). An overview of the subject is proposed in [38]. More recently, discretization schemes of higher orders (*e.g.*, $h=2$), based on cubature methods, have been introduced and studied by Kusuoka [43], Lyons [53], Ninomiya, Victoir [57] or Alfonsi [3]. The reader may also refer to the work Kohatsu-Higa and Tankov [40] for a higher weak order for jump processes. Finally, in [3], a third weak order scheme (with $h=3$) has been introduced following similar cubature ideas. This is the one we will study in this paper.

As we already precised, all those schemes converge for some $q\geqslant 1$ in (2.5). Another point of interest relies thus on the study of the set of test functions which enable the converge with weak order h. The purpose is to extend this set beyond $\mathcal{C}_b^q(\mathbb{R};\mathbb{R})$ and to obtain (2.5) with $\|f\|_{q,\infty}$ replaced by $\|f\|_\infty$ when f is a measurable and bounded function. In this case, we say that the scheme converges for the total variation distance. A first result of this type has been obtained by Bally and Talay [11], [12]. They treat the case of the Euler scheme using the Malliavin calculus (see also Guyon [34] when f is a tempered distribution). Afterwards Konakov, Menozzi and Molchanov [41], [42] established some local limit theorems using a parametrix method. Recently Kusuoka [44], also using Malliavin Calculus, obtained estimates of the error in total variation distance for the Ninomiya Victoir scheme (which corresponds to the case $h = 2$) under a Hörmander type condition.

Under an ellipticity condition, we will obtain a similar result for the case $h = 3$, using a scheme introduced in [3]. This scheme is well defined if the Lie bracket between V_1^2 and V_0 is equal to $2\tilde{V}^2$, with $\tilde{V}$ a first order differential operator. Since we consider one dimensional processes with form (2.1), we will be able to give an explicit necessary and sufficient condition in order to obtain this property on the Lie bracket.

Moreover, we will not work in a Gaussian framework and then we will have to use a variant of the Malliaivin calculus introduced by Bally and Clement [9] for which we can apply the results from [10]. A main interest of this approach is that the random variables involved in the scheme do not have a specific law but simply belong to a class of random variables which are Lebesgue lower bounded and satisfy some moment conditions. In this way, our final result can be seen as an invariance principle. The ambit of this scheme thus goes well beyond the Gaussian case.

We will begin presenting the framework of this paper in Section 2.2. In Section 2.3, we will give some third weak order convergence results for smooth test functions and for bounded measurable test functions. The latter is presented in Theorem 2.3.2 and constitutes the main result of this paper. It gives the convergence for the total variation distance with order three of the scheme from [3], toward the Markov process (2.1). We will also obtain an estimate of the density function of the diffusion and its derivatives. We will follow with a short numerical illustration in order to check the order of convergence for a suited example. This paper will end with the proof of our main theorems in Section 2.5.

2.2 The third weak order scheme

We consider the one dimensional $\mathbb{R}$-valued diffusion process

$$dX_t = V_0(X_t)dt + V_1(X_t) \circ dW_t, \tag{2.6}$$

with $V_0, V_1 \in \mathcal{C}_b^\infty(\mathbb{R};\mathbb{R})$, $(W_t)_{t\geqslant 0}$ a standard Brownian motion. Moreover, $\circ dW_t$ denotes the Stratonovich integral with respect to W. The infinitesimal operator of this Markov process is

$$A = V_0 + \frac{1}{2}V_1^2, \tag{2.7}$$

with the notation $Vf(x) = V(x)\partial f(x)$. Let us define $\exp(V)(x) := \Phi_V(x,1)$ where Φ_V solves the deterministic equation

$$\Phi_V(x,t) = x + \textstyle\int_0^t V(\Phi_V(x,s))ds. \tag{2.8}$$

By a change of variables one obtains $\Phi_{\varepsilon V}(x,t) = \Phi_V(x,\varepsilon t)$, so we have

$$\exp(\varepsilon V)(x) := \Phi_{\varepsilon V}(x,1) = \Phi_V(x,\varepsilon).$$

We also notice that the semigroup of the above Markov process, which is given by $P_t^V f(x) = f(\Phi_V(x,t))$, has the infinitesimal operator $A_V f(x) = V f(x)$. In particular the relation $P_t^V A_V = A_V P_t^V$ reads

$$Vf(\Phi_V(x,t)) = A_V P_t^V f = P_t^V A_V f = V(x)\partial_x \left(f \circ \Phi_V\right)(x,t).$$

Using m times Dynkin's formula $P_t^V f(x) = f(x) + \int_0^t P_s^V A_V f(x) ds$ we obtain

$$f(\Phi_V(x,t))) = f(x) + \sum_{r=1}^{m} \frac{t^r}{r!} V^r f(x) + \frac{1}{m!}\int_0^t (t-s)^m V^{m+1} P_s^V f(x) ds. \tag{2.9}$$

We present now the third weak order scheme introduced in [3]. In order to do it, we introduce the following commutation property:

$$V_1^2 V_0 - V_0 V_1^2 = 2\tilde{V}^2, \tag{2.10}$$

where $\tilde{V}$ is a first order operator. We consider some sequences $\epsilon_k, \rho_k,\ k \in \mathbb{N}$ of independent uniform random variables with values in $\{-1,1\}$ and $\{1,2,3\}$, and we define $\psi : \{-1,1\} \times \{1,2,3\} \times \mathbb{R}^3 \to \mathbb{R}$ using the following splitting procedure:

$$\psi(\epsilon_k, \rho_k, x, w^1_{k+1}, w^0_{k+1}) = \begin{cases} \exp(\epsilon_k w^0_{k+1}\tilde{V}) \circ \exp(w^0_{k+1}V_0) \circ \exp(w^1_{k+1}V_1)(x), & \text{if } \rho_k = 1, \\ \exp(w^0_{k+1}V_0) \circ \exp(\epsilon_k w^0_{k+1}\tilde{V}) \circ \exp(w^1_{k+1}V_1)(x), & \text{if } \rho_k = 2, \\ \exp(w^0_{k+1}V_0) \circ \exp(w^1_{k+1}V_1) \circ \exp(\epsilon_k w^0_{k+1}\tilde{V})(x), & \text{if } \rho_k = 3, \end{cases} \tag{2.11}$$

with $w^0_k = T/n$, $w^1_k = \sqrt{T} Z_k / \sqrt{n}$. We notice that $\psi(\epsilon_k, \rho_k, x, 0, 0) = x$, which is relevant with the definition of a scheme. Moreover $Z_k,\ k \in \mathbb{N}^*$ are independent random variables which are lower bounded by the Lebesgue measure: There exists $z_{*,k} \in \mathbb{R}$ and $\varepsilon_*, r_* > 0$ such that for every Borel set $A \subset \mathbb{R}$ and every $k \in \mathbb{N}^*$

$$L_{z_*}(\varepsilon_*, r_*) \qquad \mathbb{P}(Z_k \in A) \geqslant \varepsilon_* \lambda(A \cap B_{r_*}(z_{*,k})). \tag{2.12}$$

Moreover, we assume that the sequence Z_k satisfies the following moment conditions:

$$\begin{gathered} \mathbb{E}[Z_k] = \mathbb{E}[Z_k^3] = \mathbb{E}[Z_k^5] = \mathbb{E}[Z_k^7] = 0, \quad \mathbb{E}[Z_k^2] = 1, \quad \mathbb{E}[Z_k^4] = 3, \quad \mathbb{E}[Z_k^6] = 15, \\ \forall p \geqslant 1, \qquad \mathbb{E}[|Z_k|^p] < \infty. \end{gathered} \tag{2.13}$$

One step of our scheme (between times t^n_k and t^n_{k+1}) is given by

$$X^n_{t^n_{k+1}} = \psi(\epsilon_k, \rho_k, X^n_{t^n_k}, w^1_{k+1}, w^0_{k+1}). \tag{2.14}$$

Using the notation from (2.2), we also have

$$X^n_{t^n_{k+1}} = \psi_k(X^n_{t^n_k}, w^1_{k+1}, w^0_{k+1}). \tag{2.15}$$

with $\psi_k(x,z,t) = \psi(\epsilon_k, \rho_k, x, z, t)$. In the sequel, we will study the third order convergence of this scheme towards the Markov process given in (2.1) for smooth test functions and for bounded measurable test functions.

2.3 Convergence Results

We begin introducing some notations. Let $r \in \mathbb{N}^*$. For a sequence of functions $\psi_k \in \mathcal{C}^r(\mathbb{R} \times \mathbb{R} \times \mathbb{R}_+; \mathbb{R})$, $k \in \mathbb{N}$, we denote

$$\|\psi\|_{1,r,\infty} = 1 \vee \sup_{k \in \mathbb{N}} \sum_{|\alpha|=0}^{r} \sum_{|\beta|+|\gamma|=1}^{r-|\alpha|} \|\partial_x^\alpha \partial_z^\beta \partial_t^\gamma \psi_k\|_\infty, \tag{2.16}$$

and for $r \in \mathbb{N}^*$,

$$\mathfrak{K}_r(\psi) = (1 + \|\psi\|_{1,r,\infty}) \exp(\|\psi\|_{1,3,\infty}^2). \tag{2.17}$$

2.3.1 Smooth test functions

In this Scetion, we study the convergence of the scheme given in (2.15) for smooth test functions. We state a first result, which is the starting point in order to prove the convergence in total variation distance.

Theorem 2.3.1. *Suppose that $V_0, V_1, \tilde{V} \in \mathcal{C}_b^\infty(\mathbb{R}; \mathbb{R})$. We also assume that (2.10) and (2.13) hold. Then, there exists some universal constant $l \in \mathbb{N}^*$, $C \geqslant 1$ such that for every $f \in \mathcal{C}_b^8(\mathbb{R})$, we have*

$$\sup_{t \in \pi_{T,n}^T} |\mathbb{E}[f(X_t)) - \mathbb{E}[f(X_t^n)]| \leqslant C C_8(V)^l \|f\|_{8,\infty}/n^3, \tag{2.18}$$

with $C_q(V) := \sup_{i=0,1} \|V_i\|_{q,\infty} + \|\tilde{V}\|_{q,\infty}$.

Remark 2.3.1. *This result has already been obtained in [3] in the case of test functions with polynomial growth. The proof is similar and since we intend to obtain this result with the supremum norm of f we do not treat that case.*

We give a proof of this result in Section 2.5. Once we have used the Lindeberg decomposition, it relies on short time estimates using the Dynkin's formula. Now, we are going to take a step further and consider simply bounded and measurable test functions. Notice that, it means the convergence for total variation distance.

2.3.2 Bounded mesurable test functions

We see that the estimate (2.18) involves the derivatives of order eight of the test function. We will see that it is possible to obtain similar estimates with $\|f\|_{8,\infty}$ replaced by $\|f\|_\infty$. This is a consequence of a result from [10] in which the authors provide some sufficient conditions for the scheme in order to obtain the convergence for the total variation distance. The scheme (2.14) satisfy those conditions and, under an ellipticity assumption on the diffusion coefficient V_1, we are going to obtain an estimate of its total variation distance with the diffusion process (2.6).

Before doing it, we introduce a necessary and sufficient explicit condition in order to obtain (2.10) as son as for all $x \in \mathbb{R}, V_1(x) \neq 0$. Notice that, since we assume that V_1 is continuous, it has a constant sign. Moreover, this hypothesis will not be restrictive in this application. Indeed, the ellipticity condition required to use the result from [10] implies that $\inf_x V_1(x)^2 \geqslant \lambda_* > 0$

for a constant λ_*. We will suppose without loss of generality that V_1 is positive. The necessary and sufficient condition for (2.10) is the following: We assume that the function

$$\begin{aligned} g :&\mathbb{R} \to \mathbb{R} \\ &x \mapsto V_0(x)/V_1(x) \end{aligned} \tag{2.19}$$

is increasing. Notice that if V_1 is negative, g has to be decreasing.

Moreover, we propose an alternative scheme in order to approximate the density function of X and its derivatives. We consider a standard normal random variable G which is independent from $Z_k, k \in \mathbb{N}$, and for $\theta > 0$, we introduce $(X_t^{n,\theta})_{t\in\pi_{T,n}}$ as follows

$$X_t^{n,\theta}(x) = \frac{1}{n^\theta} G + X_t^n(x). \tag{2.20}$$

where $X^n(x)$ is the process which starts from x that is $X_0^n = x$. We denote by $p_t^{\theta,n}(x,y)$ the density of the law of $X_t^{\theta,n}(x)$ and for $t \in \pi_{T,n}$, we define

$$Q_t^{n,\theta} f(x) := \mathbb{E}[f(\frac{1}{n^\theta} G + X_t^n(x))]. \tag{2.21}$$

Now, we can state our main result.

Theorem 2.3.2. *Suppose that $V_0, V_1, \tilde{V} \in \mathcal{C}_b^\infty(\mathbb{R};\mathbb{R})$. We fix $T > 0$ and we also assume that (2.19), (2.12) and (2.13) hold and that*

$$V_1(x)^2 \geqslant \lambda_* > 0 \qquad \forall x \in \mathbb{R}. \tag{2.22}$$

Let $S \in (0, T/2)$. Then there exists $n_0 \in \mathbb{N}^*$ such that for every $n \geqslant n_0$, we have the following properties.

- ***A.*** *There exists $l \in \mathbb{N}^*$ and $C \geqslant 1$ which depends on m_*, r_* and the moments of Z such that, for every bounded and measurable function $f : \mathbb{R} \to \mathbb{R}$,*

$$\sup_{t\in\pi_{T,n}^{2S,T}} |\mathbb{E}[f(X_t)] - \mathbb{E}[f(X_t^n)]| \leqslant C \frac{C_8(V)^l \mathfrak{K}_{11}(\psi)^l}{(\lambda_* S)^{\eta(8)}} \left\| f \right\|_\infty / n^3. \tag{2.23}$$

  with $\mathfrak{K}_r(\psi)$ and $C_q(V)$ given in (2.17) and (2.18) and $\eta(r) = r(r+1)$.

- ***B.*** *Moreover, for every $t > 0, P_t(x,dy) = p_t(x,y)dy$ with $(x,y) \mapsto p_t(x,y)$ belonging to $\mathcal{C}^\infty(\mathbb{R}^d \times \mathbb{R}^d)$.*

- ***C.*** *Let $\theta \geqslant h+1$. We recall the $Q^{n,\theta}$ is defined in (2.21) and verifies $Q_t^{n,\theta}(x,dy) = p_t^{n,\theta}(x,y)dy$. Then, there exists $l \in \mathbb{N}^*$ such that for every $R > 0, \varepsilon \in (0,1)$, $x_0, y_0 \in \mathbb{R}^d$, and every multi-index α, β with $|\alpha| + |\beta| = u$, we also have*

$$\sup_{t\in\pi_{T,n}^{2S,T}} \sup_{(x,y)\in\overline{B}_R(x_0,y_0)} |\partial_x^\alpha \partial_y^\beta p_t(x,y) - \partial_x^\alpha \partial_y^\beta p_t^{n,\theta}(x,y)| \leqslant C \frac{C_8(V)^l \mathfrak{K}_{11}(\psi)^l}{(\lambda_* S)^{\eta(p_{u,\varepsilon} \vee 8)}} / n^{3(1-\varepsilon)} \tag{2.24}$$

  with a constant C which depends on R, x_0, y_0, T and on $|\alpha| + |\beta|$ and $p_{u,\varepsilon} = (u + 2d + 1 + 2\lceil(1-\varepsilon)(u+d)/(2\varepsilon)\rceil)$.

Remark 2.3.2. *It is relevant to notice that we have the same result if we assume that the function defined in (2.19) is decreasing (resp. increasing) for V_1 positive (resp. V_1 negative). In this case $V_0V_1^2 - V_1^2V_0 = 2\tilde{V}^2$ and we have to define the scheme differently. In the construction (2.11), we invert the terms containing V_1 with the ones containing V_0.*

Remark 2.3.3. *The property (2.12) is crucial here, since we will use a result from [10] which employs abstract integration by parts formulae based on the noise Z_k. However it is not restrictive for concrete applications.*

The result (2.23) signifies the convergence in total variation with order 3. The proof of this theorem is given in Section 2.5. Since we have already obtained some short time estimates of the form (2.4) in the proof of Thereom 2.3.1 and (2.19) holds, the key point of this proof does not rely on the weak order of the scheme. This is the fact that, the splitting procedure (2.11) in order to build the scheme, always includes a diffusion part through $\exp(Z_k/\sqrt{n/T}V_1)$, with Z_k satisfying (2.12) and the ellipticity condition (2.22) for V_1. The proof is then a consequence from Theorem 3.3 in [10] which employs an abstract Malliavin calculus based on such noise Z_k and initially presented by Bally and Clément [9]. A similar approach can be used in order to prove the convergence for the total variation distance for even higher order scheme built as in (2.11). The main difficulty will then rely on the proof of the short time estimate (2.4).

A main interest of this result is that it can be seen as an invariance principle as well. Indeed, it does not require that Z_k follows a particular law but only the properties (2.12) and (2.13). In particular, we do not restrict ourselves to the Gaussian framework which is necessary to use the Malliavin Calculus in order to proove the convergence for the total variation distance as in [11], [12], or [44]. In this way, the condition (2.12) might be a hint to find a necessary condition on the random variables $(Z_k)_{k\in\mathbb{N}^*}$ in order to obtain the total variation convergence with order $h = 3$.

Moreover, using Remark 2.3.2, we can define a third order scheme as soon as the function defined in (2.19) is monotonic. If it is increasing (recall that $V_1(x) \geqslant 0$), the Lie bracket between V_1^2 and V_0 is given by $[V_1^2, V_0]f = V_1^2V_0f - V_0V_1^2f = 2\tilde{V}^2f \quad \text{with}$

$$\tilde{V}(x) = \sqrt{|V_1(x)(V_1(x)\partial_xV_0(x) - \partial_xV_1(x)V_0(x))|}, \quad x \in \mathbb{R}.$$

If it is decreasing, we have $[V_0, V_1^2]f = 2\tilde{V}^2f$ as well. This explicit representation for $\tilde{V}$ is crucial for concrete applications since the scheme is defined using the solution of (2.8) with $V = \tilde{V}$. Moreover, looking at (2.18) and (2.23), we have to control its derivatives.

2.4 Numerical illustration

In this section, we study the numerical approximation of a one dimensional SDE with schemes defined on homogeneous time grids with form $\pi_{T,n} = \{kT/n, k \in \mathbb{N}\}$. We will fix T and we will analyze the behavior of the total variation distance between the diffusion process $(X_t)_{t\geqslant 0}$ and miscellaneous discretization schemes $(X_t^n)_{t\in\pi_{T,n}}$ with respect to the number of time step n. More particularly, we will study the weak error $|\mathbb{E}[f(X_t)] - \mathbb{E}[f(X_t^n)]|$ for bounded measurable functions f and various n.

In concrete applications, once we have selected a scheme X^n, $\mathbb{E}[f(X_t^n)]$ will be used to estimate $\mathbb{E}[f(X_t)]$. The next step is thus to approximate $\mathbb{E}[f(X_t^n)]$. A standard way to do it, is to use a Monte Carlo method. Given an independent sampling of size M, and using the Central Limit

Theorem, we can easily show that those algorithms converge toward the real expectancy with rate $\sqrt{M}$. Moreover, discretization schemes provide an estimation of $\mathbb{E}[f(X_t)]$ with any desired precision since we can choose any value for n. However, the cost of calculation will also increase with n since we have n iterations of the scheme function (2.2). At this point, it is important to notice that there is a trade off to make between the precision we want to obtain and the time of calculation we can afford. Indeed, if our scheme converges with order h, we have to choose $M = \mathcal{O}(n^{2h})$ and then choose n large enough in order to obtain the desired precision. We will see that even if the time of calculation of one step of the scheme we study in this paper is much longer than the time of a lower order scheme (*e.g.* the Euler scheme), the third weak order scheme is better in time of calculation and precision as soon as the precision is high enough. In order to illustrate the reason why we point out such properties, we now present our example.

We consider the Markov diffusion process $(X_t)_{t\geqslant 0}$ given by the following SDE,

$$dX_t = adt + \frac{\sigma}{\arctan(X_t) + \pi} \circ dW_t, \tag{2.25}$$

with $\sigma > 0$ and $a \in \mathbb{R}$. Notice that the coefficients of the SDE (2.25) belong to $\mathcal{C}_b^\infty(\mathbb{R})$ and moreover $V_1 : x \mapsto \sigma/(\arctan(x) + \pi)$ satisfies $\inf_x V_1(x) > 2\sigma/\pi$ and the function $V_0/V_1 = a/V_1$ is increasing. Therefore, the scheme (2.11) is well-defined and we have the required hypothesis in order to obtain the results from Theorem 2.3.2. Moreover, we have an explicit representation for the first order operator $\tilde{V}$, that is : $\tilde{V}(x) = \sigma\sqrt{a}/(\sqrt{1+x^2}(\arctan(x)+\pi)^{3/2})$.

The next step consists then in solving the ODE (2.8) for $V = V_0, V_1, \tilde{V}$. Looking closer to (2.11), we will use each of these solutions once for each step of the discretization algorithm. In this example, it is easy to find an analytic solution to (2.8) when $V = V_0$. However for $V = V_1, \tilde{V}$, it is much more cumbersome and we will use some numerical algorithms. A naive algorithm consists in using the Riemann approximation of $\int_0^t V(\Phi_V(x,s))ds$ on a time grid of $[0,t]$ in the following way : For a number N of time steps, we put $\Phi_V^N(x,0) = x$ and for $i \in \{0,\ldots,N-1\}$, $\Phi_V^N(x,(i+1)t/N) = \Phi_V^N(x,it/N) + TN^{-1}V(\Phi_V^N(x,it/N))$. This is the method we will use to approximate $\Phi_V^N(x,t)$. Finally, in this case, we can use an alternative way to approximate $\Phi_{V_1}^N(x,t)$. Indeed, we can show that $g(\Phi_{V_1}^N(x,t)) = g(x) + t$ where g is the bijective function in $\mathcal{C}^1(\mathbb{R})$ defined by

$$g(x) = (x\arctan(x) - 0.5\log(1+x^2) + x\pi)/\sigma.$$

Then, we can find $\Phi_{V_1}(x,t)$ using a Newton algorithm in order to invert g. Likewise the naive Riemann approximation, this method provides an approximation given a parameter of precision (which is N for Riemann sums). Obviously, the more this parameter is tight, the more the cost of the algorithm is high. Compared to one step the Euler scheme,

$$\begin{aligned} X_{t_{k+1}^n}^{n,Eul} &= X_{t_k^n}^{n,Eul} + (a - \frac{\sigma^2}{2(1+(X_{t_k^n}^{n,Eul})^2)(\arctan(X_{t_k^n}^{n,Eul})+\pi)^3})T/n \\ &+ \frac{\sigma}{\arctan(X_{t_k^n}^{n,Eul})+\pi}\sqrt{T/N}Z_{k+1}, \quad (Z_k)_{k\in\mathbb{N}^*}\ i.i.d \sim \mathcal{N}(0,1), \end{aligned} \tag{2.26}$$

the cost of one step of (2.11) can thus be very important. However, despite that cost, the third order scheme become more effective as soon as we want to compute $\mathbb{E}[f(X_t)]$ with a sufficiently high precision.

Heuristically, let $\epsilon > 0$, the precision of the weak error that is $|\mathbb{E}[f(X_t)] - \mathbb{E}[f(X_t^n)]| \leqslant \epsilon$. In order to reach that precision, we will have to run $M = \epsilon^{-2}$ Monte Carlo iterations. Now let

$n \in \mathbb{N}$ such that $n^3 = \epsilon^{-1}$. Then, if we want to reach this precision, we will have to simulate $M = \epsilon^{-2}$ realizations of the third order scheme with time step t/n, or of the Euler scheme with time step t/n^3. Now, we assume that the cost in time of calculation of one step of the third order scheme is given by τ_{NV3} and by τ_{Eul} for the Euler scheme. Then the total cost to reach the precision ϵ will be $\tau_{NV3} n M = \tau_{NV3}\epsilon^{-2-1/3}$ for the third order scheme and $\tau_{Eul} n^3 M = \tau_{Eul}\epsilon^{-3}$ for the Euler scheme. Then, as soon as $\tau_{NV3}/\tau_{Eul} \leqslant \epsilon^{-2/3}$, the cost of the third order scheme will be lower than the cost of the Euler scheme. Controversially, if τ_{NV3} and τ_{Eul} are fixed we can find a precision ϵ_0 such that the cost of the three order scheme is lower than the one of the Euler scheme for all $\epsilon \leqslant \epsilon_0$.

In Figure 2.1, we represent the error $|\mathbb{E}[f(X_t)] - \mathbb{E}[f(X_t^n)]|$ [1], with respect to the number of time steps n, in Log Log scale, for the third order scheme we study in this paper and when f is a Heavyside function. We observe that the scheme converges with the expected rate, that is $h \approx 2.91$. This numerical experiment thus confirms the total variation convergence result from Theorem 2.3.2. Notice that we have also implemented the Euler scheme and the Ninomiya Victoir scheme of order 2 [57] in order to compare the cost of the different approaches. With the precision parameters we have selected in the algorithms solving (2.8) in order to obtain Figure 2.1, we have $\tau_{NV3} \approx 7.8\tau_{NV2} \approx 51.9\tau_{Eul}$ which is quite reasonable given the gain which is made with respect to the number of time steps. In this case, the third order scheme thus become more effective than the Euler scheme as soon as the precision ϵ of the weak error satisfies $|\mathbb{E}[f(X_t)] - \mathbb{E}[f(X_t^n)]| \leqslant \epsilon \leqslant (51.9)^{-3/2}$.

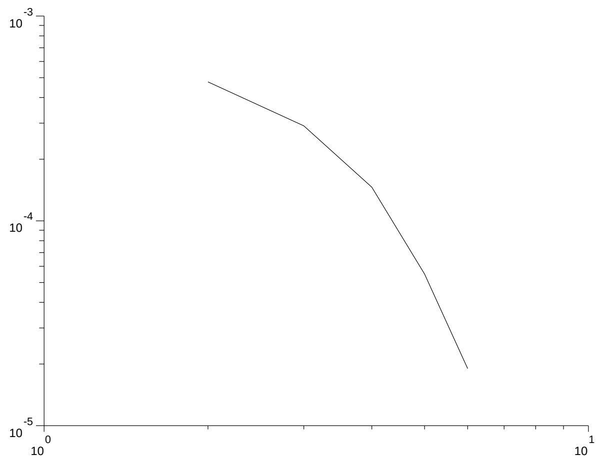


Figure 2.1: Log-Log representation of $|\mathbb{E}[f(X_t)] - \mathbb{E}[f(X_t^n)]|$ for $x = 0.8$, $T = 1$, $a = 0.2$, $\sigma = 2$, with respect to n for $f(x) = \mathbb{1}_{x>1.1}$.

[1] We do not estimate $\mathbb{E}[f(X_t)]$ using Monte Carlo methods with exact simulation but with the third order scheme for $n = 50$.

2.5 Proof of the main theorems

Proof of Theorem 2.3.1. **Step 1.** We define $(P_{t,s})_{t,s\in\pi_{T,n};t\leqslant s}$ by

$$P^n_{t,t}f(x)=f(x),\qquad \forall k\leqslant r\in\mathbb{N},\ P^n_{t^n_k,t^r_k}f=\mathbb{E}[f(X_{t^n_r})|X_{t^n_k}=x],$$
$$Q^n_{t,t}f(x)=f(x),\qquad \forall k\leqslant r\in\mathbb{N},\ Q^n_{t^n_k,t^r_k}f=\mathbb{E}[f(X^n_{t^n_r})|X^n_{t^n_k}=x],$$

and we notice that for $t,s,u\in\pi_{T,n}$ with $t\leqslant s\leqslant u$, then $P^n_{t,u}f=P^n_{t,s}P^n_{s,u}f$. It follows that

$$\begin{aligned}\Big|\mathbb{E}[f(X_{t^n_m})]-\mathbb{E}[f(X^n_{t^n_m})]\Big| &\leqslant \|P^n_{0,t^n_m}f-Q^n_{0,t^n_m}f\|_\infty \\ &\leqslant \sum_{k=0}^{m-1}\|P^n_{t^n_{k+1},t^n_m}P^n_{t^n_k,t^n_{k+1}}Q_{t^n_{k+1}}f-P^n_{t^n_{k+1},t^n_m}Q_{t^n_k,t^n_{k+1}}Q^n_{t^n_k}f\|_\infty \\ &=\sum_{k=0}^{m-1}\|P_{t^n_{k+1},t^n_m}(P_{t^n_k,t^n_{k+1}}-Q_{t^n_k,t^n_{k+1}})Q_{t^n_k}f\|_\infty.\end{aligned}\tag{2.27}$$

$$\begin{aligned}\|P^n_{t^n_m}f-Q^n_{t^n_m}f\|_\infty &\leqslant \sum_{k=0}^{m-1}\|P^n_{t^n_k}P^n_{t^n_k,t^n_{k+1}}Q^n_{t^n_{k+1},t^n_m}f-P^n_{t^n_k}Q^n_{t^n_k,t^n_{k+1}}Q^n_{t^n_{k+1},t^n_m}f\|_\infty \\ &=\sum_{k=0}^{m-1}\|P^n_{t^n_k}(P_{t^n_k,t^n_{k+1}}-Q_{t^n_k,t^n_{k+1}})Q^n_{t^n_{k+1},t^n_m}f\|_\infty.\end{aligned}\tag{2.28}$$

We notice that it easy to prove that, for $t,s\in\pi_{T,n}$, $t\leqslant s$, $\|P_{t,s}f\|_{p,\infty}\leqslant C\|f\|_{p,\infty}$ and $\|Q_{t,s}f\|_{p,\infty}\leqslant C\|f\|_{p,\infty}$.
Step 2. It remains to show $\|P_{t^n_k,t^n_{k+1}}f-Q_{t^n_k,t^n_{k+1}}f\|_\infty\leqslant C\|f\|_{8,\infty}/n^4$ and using (2.28) the proof will be completed. In order to simplify the notation, we fix $T=1$ without loss of generality. For $\epsilon=-1,1$, we denote

$$\mathcal{T}_0f(x)=f(\exp(\frac{1}{n}V_0)(x)),\quad \mathcal{T}_1f(x)=f(\exp(\frac{Z}{\sqrt{n}}V_1)(x)),\quad \tilde{\mathcal{T}}_\epsilon f(x)=f(\exp(\frac{\epsilon}{n}\tilde{V})(x)).$$

Notice that, using the notation introduced in the beginning of this section with $V=n^{-1/2}ZV_1$, we have $\mathcal{T}_1f(x)=P_1^{n^{-1/2}ZV_1}f(x)$. Using (2.9) with $t=1$ and $V=n^{-1/2}ZV_1$ we obtain

$$\mathcal{T}_1f(x)=f(x)+\sum_{r=1}^m\frac{Z^r}{n^{r/2}}\frac{1}{r!}V_1^rf(x)+\frac{Z^{m+1}}{n^{(m+1)/2}}R_{m+1,1}f(x)\tag{2.29}$$

with

$$R_{m+1,1}f(x)=\frac{1}{m!}\int_0^1(1-\lambda)^mV_1^{m+1}P_\lambda^{n^{-1/2}ZV_1}f(x)d\lambda\tag{2.30}$$

and we recall that $P_\lambda^{n^{-1/2}ZV_1}f(x)=f(\exp(\lambda ZV_1/\sqrt{n}))$. We have a similar development if we put $V=V_0/n$ or $V=\epsilon\tilde{V}/n$ in (2.9). Our aim is to give a development of order 4 (with respect to n) for $\mathbb{E}[f(\psi_k(x,w^1_{k+1},w^0_{k+1})]$ (see (2.31) below). We replace each $\mathcal{T}\in\{\mathcal{T}_0,\mathcal{T}_1,\tilde{\mathcal{T}}_\epsilon\}$, with an expansion of order $m\leqslant 7$ given above with $Z=Z_{k+1}$ for $\mathcal{T}_1$ and $m\leqslant 3$ for $\mathcal{T}=\mathcal{T}_0,\tilde{\mathcal{T}}$. Then, we calculate the products of the miscellaneous expansions, each with a well chosen order such that there is no term with factor n^{-r}, $r>4$, appearing in those productss. Moreover, all the terms containing n^{-4} go in the remainder. The last step consists in computing the expectancy. We notice that $\mathbb{E}[P_t^{n^{-1/2}ZV_1}] = P_t^{V_1^2/(2n)}$ and $\mathbb{E}[Z_{k+1}^r] = 0$ for odd $r \leqslant 7$. Finally, since $\mathbb{E}[Z_{k+1}^2] = 1$, $\mathbb{E}[Z_{k+1}^4] = 6$ and $\mathbb{E}[Z_{k+1}^6] = 15$, the calculus is completed and we obtain:

$$\mathbb{E}[f(\psi_k(x, w_{k+1}^1, w_{k+1}^0)] = \frac{1}{6}\sum_{\epsilon=-1,1} \mathbb{E}[(\tilde{\mathcal{T}}_\epsilon \mathcal{T}_0 \mathcal{T}_1 + \mathcal{T}_0 \tilde{\mathcal{T}}_\epsilon \mathcal{T}_1 + \mathcal{T}_0 \mathcal{T}_1 \tilde{\mathcal{T}}_\epsilon) f(x)] \tag{2.31}$$

$$\begin{aligned}
&= f(x) + \frac{1}{n}(V_0 + \frac{1}{2}V_1^2)f(x) + \frac{1}{2n^2}(V_0^2 + \frac{1}{4}V_1^4 + 2V_0\frac{1}{2}V_1^2 + \tilde{V}^2)f(x) \\
&\quad + \frac{1}{6n^3}(\frac{1}{8}V_1^6 + V_0^3 + 3V_0\frac{1}{4}V_1^4 + 3V_0^2\frac{1}{2}V_1^2 + 2\tilde{V}^2\frac{1}{2}V_1 + 2V_0\tilde{V}^2 + \frac{1}{2}V_1^2\tilde{V}^2 + \tilde{V}^2 V_0)f(x) \\
&\quad + \frac{1}{n^4}Rf(x) \\
&= f(x) + \frac{1}{n}(V_0 + \frac{1}{2}V_1^2)f(x) + \frac{1}{2n^2}(V_0 + \frac{1}{2}V_1^2)^2 f(x) + \frac{1}{6n^3}(V_0 + \frac{1}{2}V_1^2)^3 f(x) + \frac{1}{n^4}Rf(x)
\end{aligned}$$

The remainder R is a sum of terms of the following form:

$$C(\tilde{\mathcal{T}}_{\epsilon,\alpha_\epsilon}\mathcal{T}_{0,\alpha_0}\mathcal{T}_{1,\alpha_1} + \mathcal{T}_{0,\alpha_0}\tilde{\mathcal{T}}_{\epsilon,\alpha_\epsilon}\mathcal{T}_{1,\alpha_1} + \mathcal{T}_{0,\alpha_0}\mathcal{T}_1\tilde{\mathcal{T}}_{\epsilon,\alpha_\epsilon})f(x) \tag{2.32}$$

with $\alpha = (\alpha_0, \alpha_1, \alpha_2) \in \{0, \ldots, 4\}^3$, $|\alpha| = \alpha_0 + \alpha_1 + \alpha_2 = 4$, and, using the notation given in (2.30),

$$\begin{aligned}
\mathcal{T}_{0,k} &\in \{V_0^k, R_{k,0}\}, & \tilde{\mathcal{T}}_{\epsilon,k} &\in \{\tilde{V}^k, \tilde{R}_{k,\epsilon}\}, & \mathcal{T}_{1,k} &\in \{V_1^{2k}, R_{2k,1}\}, \quad k = 0, \ldots, 3, \\
\mathcal{T}_{0,4} &= R_{4,0}, & \tilde{\mathcal{T}}_{\epsilon,4} &= \tilde{R}_{4,\epsilon}, & \mathcal{T}_{1,4} &= \mathbf{R}_{8,1},
\end{aligned}$$

with

$$\mathbf{R}_{8,1} = \mathbb{E}[Z^8 R_{8,1}] = \int_0^1 (1-\lambda)^7 \mathbb{E}[Z^8 V_1^8 P_\lambda^{U_1} f(x)] d\lambda.$$

It is easy to check that for every $g \in \mathcal{C}^{k+p}(\mathbb{R})$, we have the following property

$$\|\mathcal{T}_{0,k}g\|_{p,\infty} + \|\mathcal{T}_{1,k}g\|_{p,\infty} + \|\tilde{\mathcal{T}}_{\epsilon,k}g\|_{p,\infty} \leqslant CC_{2k+p}(V)^l \|g\|_{2k+p,\infty}$$

for some constants $l \in \mathbb{N}^*$, $C \geqslant 1$. So

$$\|Rf\|_\infty \leqslant CC_8(V)^l \|f\|_{8,\infty}. \tag{2.33}$$

We turn now to the diffusion process X_t. We have the development

$$\mathbb{E}[f(X_t(x))] = P_t^A f(x) = f(x) + tAf(x) + \frac{t^2}{2}A^2 f(x) + \frac{t^3}{6}A^3(x) + \frac{t^4}{4!}R_t' f(x).$$

with

$$R_t' f(x) = t^{-1}\int_0^t P_\lambda^A A^4 f(x)(1-\lambda/t)^3 d\lambda. \tag{2.34}$$

We take $t = n^{-1}$ and make the difference between (2.34) and (2.31). All the terms cancel except for the remainders so we obtain

$$\begin{aligned}
&\forall k \in \{0, \ldots, n-1\}, \\
&\qquad \mathbb{E}[f(X_{t_{k+1}^n})] - \mathbb{E}[f(X_{t_{k+1}^n}^n) \mid X_{t_k^n} = X_{t_k^n}^n = x] = (R_{1/n}' f(x)/4! - Rf(x))/n^4.
\end{aligned} \tag{2.35}$$

We clearly have $\|R_{1/n}' f\|_\infty \leqslant CC_8(V)^l \|f\|_{8,\infty}$. This, together with (2.33) and (2.28), completes the proof. □

Proof of Theorem 2.3.2. **Step 1.** Let us prove that (2.10) is satisfied. We have

$$\begin{aligned}\frac{1}{2}(V_1^2V_0 - V_0V_1^2)f(x) =&(\partial_x V_1(x)(V_1(x)\partial_x V_0(x) - \partial_x V_1(x)V_0(x))\\ &+V_1(x)(\partial_x^2 V_0(x)V_1(x) - \partial_x^2 V_1(x)V_0(x)))\partial_x f(x)\\ &+V_1(x)(V_1(x)\partial_x V_0(x) - \partial_x V_1(x)V_0(x))\partial_x^2 f(x).\end{aligned}$$

Since $V_1(x) \neq 0$, if we take

$$\tilde{V}(x) = \sqrt{V_1(x)(V_1(x)\partial_x V_0(x) - \partial_x V_1(x)V_0(x))}, \tag{2.36}$$

then, using (2.19), $\tilde{V}$ is well defined and satisfies (2.10).

Step 2. Now we are going to show the convergence in total variation distance. In order to do it we will use a result from [10]. First, applying the same reasoning as in the proof of Theorem 2.3.1 we can show that there exists some universal constants $C, l \geqslant 1$ such that

$$|\langle g, P^n_{t_k^n,t_{k+1}^n} f - Q_{t_k^n,t_{k+1}^n} f\rangle| \leqslant n^{-4} C C_8(V)^l \|g\|_{1,8} \|f\|_\infty, \tag{2.37}$$

with $\langle \cdot,\cdot\rangle$ the scalar product in $\mathrm{L}^2(\mathbb{R})$. Now we have (2.35) and (2.37), the result will be a consequence of Theorem 3.3. in [10], as soon as we check that the following ellipticity assumption holds:

$$\exists \lambda_* > 0, \qquad \inf_{k\leqslant n}\inf_{x\in\mathbb{R}} (\partial_{w^1}\psi_k(x,w^1,w^0)_{|w^1=w^0=0})^2 \geqslant \lambda_*. \tag{2.38}$$

We fix k and we look at $\psi_k(x,w^1,w^0)$ defined in (2.15). We suppose that $\rho_k = 3, \epsilon_k = 1$ (the proof for $\rho_k = 1,2$ or $\epsilon_k = -1$ is similar). We consider the process $x_t(\tilde{w}), 0 \leqslant t \leqslant T_3$, with $T_i = i$, $\tilde{w} = (w^1, w^0)$, solution of the following equation:

$$\begin{aligned} x_t(\tilde{w}) &= x + w^0\int_0^t \tilde{V}(x_s(\tilde{w}))ds, && T_0 \leqslant t \leqslant T_1,\\ x_t(\tilde{w}) &= x_{T_1}(\tilde{w}) + w^1\int_{T_1}^t V_1(x_s(\tilde{w}))ds, && T_1 \leqslant t \leqslant T_2,\\ x_t(\tilde{w}) &= x_{T_2}(\tilde{w}) + w^0\int_{T_2}^t V_0(x_s(\tilde{w}))ds, && T_2 \leqslant t \leqslant T_3.\end{aligned}$$

We notice that $\psi_k(x,w^1_{k+1},w^0_{k+1}) = x_{T_3}(\tilde{w}_{k+1})$ and consequently, we have $\partial_z\psi_k(x,w^1_{k+1},w^0_{k+1}) = \partial_{w^1}x_{T_3}(\tilde{w}_{k+1})$. Moreover, $\partial_{w^1}x_t(w) = 0$ for $t \leqslant T_1$. Now, let $T_1 \leqslant t \leqslant T_2$. Then $\partial_{w^1}x_t(\tilde{w})$ solves the equation

$$\partial_{w^1}x_t(\tilde{w}) = \int_{T_1}^t V_1(x_s(\tilde{w}))ds + w^1\int_{T_1}^t \partial V_1(x_s(\tilde{w}))\partial_{w^1}x_s(\tilde{w})ds.$$

It follows that

$$\partial_{w^1}x_t(\tilde{w}) \mid_{\tilde{w}=0} = \int_{T_1}^t V_1(x_s(0))ds = V_1(x)(t-T_1).$$

Notice that $T_2 - T_1 = 1$. Then, we have

$$\partial_{w^1}x_{T_3}(\tilde{w}) \mid_{w^1=0} = \partial_{w^1}x_{T_2}(\tilde{w}) \mid_{\tilde{w}=0} = V_1(x).$$

and then, by (2.22),

$$(\partial_{w^1}x_{T_3}(0))^2 \geqslant \lambda_*.$$

□

Part III

Estimation of the parameters of the Wishart processes

Chapter 1

Maximum likelihood estimation for Wishart processes

Ce Chapitre est un article écrit avec A.Alfonsi et A.Kebaier actuellement en soumission [4].

Abstract

In the last decade, there has been a growing interest to use Wishart processes for modelling, especially for financial applications. However, there are still few studies on the estimation of its parameters. Here, we study the Maximum Likelihood Estimator (MLE) in order to estimate the drift parameters of a Wishart process. It turns out that this estimator is only well defined when the matrix parameter in the drift is symmetric. We obtain precise convergence rates and limits for this estimator in the ergodic case and in some nonergodic cases, that are different in each studied case. We also check that it achieves the optimal convergence rate. Motivated by this study, we also present new results on the Laplace transform that extend the recent findings of Gnoatto and Grasselli [29] and are of independent interest.

1.1 Introduction and preliminary results

The goal of this paper is to study the maximum likelihood estimation of the parameters of Wishart processes. These processes have been introduced by Bru [20] and take values in the set of positive semidefinite matrices. Let $d \in \mathbb{N}^*$ denote the dimension, $\mathcal{M}_d$ be the set of real d-square matrices, $\mathcal{S}_d^+$ (resp. $\mathcal{S}_d^{+,*}$) be the subset of positive semidefinite (resp. definite) matrices, $\mathcal{S}_d$ (resp. $\mathcal{A}_d$) the subset of symmetric (resp. antisymmetric) matrices. Wishart processes are defined by the following SDE

$$\begin{cases} dX_t = \left[\alpha a^\top a + bX_t + X_t b^\top\right] dt + \sqrt{X_t} dW_t a + a^\top dW_t^\top \sqrt{X_t}, \quad t > 0 \\ X_0 = x \in \mathcal{S}_d^+, \end{cases} \tag{1.1}$$

where $\alpha \geqslant d-1$, $a \in \mathcal{M}_d$, $b \in \mathcal{M}_d$ and $(W_t)_{t\geqslant 0}$ denotes a d-square matrix made of independent Brownian motions. We recall that for $x \in \mathcal{S}_d^+$, $\sqrt{x}$ is the unique matrix in $\mathcal{S}_d^+$ such that $\sqrt{x}^2 = x$. It is shown by Bru [20] and Cuchiero et al. [21] in a more general affine setting that the SDE Equation (1.1) has a unique strong solution when $\alpha \geqslant d+1$ and a unique weak solution when $\alpha \geqslant d-1$. Besides, we have $X_t \in \mathcal{S}_d^{+,*}$ for any $t \geqslant 0$ when $x \in \mathcal{S}_d^{+,*}$ and $\alpha \geqslant d+1$. In this paper, we will denote by $WIS_d(x, \alpha, b, a)$ the law of $(X_t, t \geqslant 0)$ and $WIS_d(x, \alpha, b, a; t)$ the law of X_t. In dimension $d = 1$, Wishart processes are known as Cox-Ingersoll-Ross processes in the literature. It is worth recalling that the law of X only depends on a through $a^\top a$ since we have

$$WIS_d(x,\alpha,b,a) \underset{law}{=} WIS_d(x,\alpha,b,\sqrt{a^\top a}),$$

see e.g. equation (12) in [1]. Therefore, the parameters to estimate are α, b and $a^\top a$.
Wishart processes have been originally considered by Bru [19] to model some biological data. Recently, they have been widely used in financial models in order to describe the evolution of the dependence between assets. Namely, Gourieroux and Sufana [32] and Da Fonseca et al. [23] have proposed a stochastic volatility model for a basket of assets that assumes that the instantaneous covariance between the assets follows a Wishart process. This extends the well-known Heston model [35] to many assets. Wishart processes have also been used for interest rates models. Affine term structure models involving these processes have been proposed for example by Gourieroux and Sufana [33], Gnoatto [28] and Ahdida et al. [2]. For these models, the question of estimating the parameters of the underlying Wishart process may be important for practical purposes and should be possible thanks to the profusion of financial data. This issue has been considered by Da Fonseca et al. [22] for the model presented in [23]. However, there is no dedicated study on the Maximum Likelihood Estimator (MLE) for Wishart processes, and the MLE has been very recently studied for the Cox-Ingersoll-Ross process by Ben Alaya and Kebaier [13, 14]. This paper completes the literature by studying the MLE for Wishart processes.
In this paper, we will follow the theory developed in the books by Lipster and Shiryaev [50] and Kutoyants [45] and assume that we observe the full path $(X_t, t\in[0,T])$ up to time $T>0$. This choice will be convenient from a mathematical point of view to study the convergence of the MLE. Of course, in practice it can be relevant to study precisely the estimation when we only observe the process on a discrete time-grid. This is left for further research, but we already observe in our numerical experiments that the discrete approximation of the MLE gives a satisfactory estimation of Wishart parameters (see Section 1.6). It is worth noticing that once we observe the path $(X_t, t\in[0,T])$, the parameter $a^\top a$ is known. In fact, we can calculate the quadratic covariation (see for example Lemma 2 in [1]) and get for $i,j,k,l\in\{1,\cdots,d\}$

$$\langle X_{i,j}, X_{k,l}\rangle_T = \int_0^T (a^\top a)_{j,l}(X_s)_{i,k} + (a^\top a)_{j,k}(X_s)_{i,l} + (a^\top a)_{i,l}(X_s)_{j,k} + (a^\top a)_{i,k}(X_s)_{j,l} ds. \quad (1.2)$$

This leads to

$$\begin{aligned}(a^\top a)_{i,i} &= \frac{1}{4}\langle X_{i,i}\rangle_T\Big(\int_0^T (X_s)_{i,i}ds\Big)^{-1}, \quad (1.3)\\ (a^\top a)_{i,j} &= \left(\frac{1}{2}\langle X_{i,j}, X_{i,i}\rangle_T - (a^\top a)_{i,i}\int_0^T (X_s)_{i,j}ds\right)\Big(\int_0^T (X_s)_{i,i}ds\Big)^{-1},\end{aligned}$$

for $1\leqslant i,j\leqslant d$ and $j\neq i$. We note that these quantities are well defined as soon as the path $(X_t, t\in[0,T])$ has a finite quadratic variation and is such that $X_t\in\mathcal{S}_d^{+,*}$ dt-a.e., which is satisfied by the paths of Wishart processes (see Proposition 4 in [20]). We will assume that $a^\top a\in\mathcal{S}_d^{+,*}$ and denote by $a\in\mathcal{M}_d$ an invertible matrix that matches the observed value of $a^\top a$: a can be for example the square root of $a^\top a$ or the Cholesky decomposition of $a^\top a$. Then, we know that $Y_t=(a^\top)^{-1}X_t a^{-1}$ follows the law $WIS_d((a^\top)^{-1}xa^{-1},\alpha,(a^\top)^{-1}ba^\top, I_d)$, see e.g. equation (13) in [1]. It is therefore sufficient to focus on the estimation of the parameters α and b when $a=I_d$, which we consider now.
We first present the MLE of $\theta=(\alpha,b)$, and we denote by $\mathbb{P}_\theta$ the original probability measure under which X satisfies

$$dX_t = \Big[\alpha I_d + bX_t + X_t b^\top\Big]dt + \sqrt{X_t}dW_t + dW_t^\top\sqrt{X_t}. \quad (1.4)$$

When no confusion is possible, we also denote $\mathbb{P}$ this probability. We consider $\alpha_0 \geqslant d+1$ and set $\theta_0 = (\alpha_0, 0)$. We will assume for the joint estimation of α and b that

$$\alpha \geqslant d+1 \text{ and } x \in \mathcal{S}_d^{+,*}. \tag{1.5}$$

The latter assumption is not restrictive in practice since the condition $\alpha \geqslant d+1$ ensures that $X_t \in \mathcal{S}_d^{+,*}$ for any $t > 0$. Due to this assumption, we know by Theorem 4.1 in Mayerhofer [54] that

$$\frac{d\mathbb{P}_{\theta_0,T}}{d\mathbb{P}_{\theta,T}} := \exp\left(\int_0^T \mathrm{Tr}[H_s dW_s] - \frac{1}{2}\int_0^T \mathrm{Tr}[H_s H_s^\top] ds\right), \text{ with } H_t = \frac{\alpha_0 - \alpha}{2}(\sqrt{X_t})^{-1} - b\sqrt{X_t}$$

defines a probability measure under which $\tilde{W}_t = W_t - \int_0^t H_s^\top ds$ is a $d \times d$-Brownian motion, where $\mathbb{P}_{\theta,T}$ is the restriction of $\mathbb{P}_\theta$ to the σ-algebra $\sigma(W_s, s \in [0,T])$. We have

$$dX_t = \alpha_0 I_d dt + \sqrt{X}_t d\tilde{W}_t + d\tilde{W}_t^\top \sqrt{X}_t,$$

and therefore X follows a Wishart process with parameter θ_0 under $\mathbb{P}_{\theta_0}$. Conversely, still by Theorem 4.1 in [54] we know that

$$\begin{aligned}
\frac{d\mathbb{P}_{\theta,T}}{d\mathbb{P}_{\theta_0,T}} &= \exp\left(-\int_0^T \mathrm{Tr}[H_s d\tilde{W}_s] - \frac{1}{2}\int_0^T \mathrm{Tr}[H_s H_s^\top] ds\right) \\
&= \exp\left(\frac{\alpha - \alpha_0}{2}\int_0^T \mathrm{Tr}[(\sqrt{X_s})^{-1} d\tilde{W}_s] + \int_0^T \mathrm{Tr}[b\sqrt{X_s} d\tilde{W}_s]\right. \\
&\qquad \left. - \frac{(\alpha - \alpha_0)^2}{8}\int_0^T \mathrm{Tr}[X_s^{-1}] ds - \frac{1}{2}\int_0^T \mathrm{Tr}[b X_s b^\top] ds - \frac{(\alpha-\alpha_0)T}{2}\mathrm{Tr}[b]\right).
\end{aligned} \tag{1.6}$$

is also a change of probability, and the probability measures $\mathbb{P}_{\theta,T}$ and $\mathbb{P}_{\theta_0,T}$ are equivalent. To see the exponential in Equation (1.6) as the likelihood of the path $(X_t, t \in [0,T])$, one has to write it as a function of $(X_t, t \in [0,T])$. This is however not possible in general unless b is a symmetric matrix.

Proposition 1.1.1. *Let $(\mathcal{F}_t^X)_{t \geqslant 0}$ denote the filtration generated by the process X. Then, $\frac{d\mathbb{P}_{\theta,T}}{d\mathbb{P}_{\theta_0,T}} \in \mathcal{F}_T^X \iff b \in \mathcal{S}_d$, and we have in this case $\frac{d\mathbb{P}_{\theta,T}}{d\mathbb{P}_{\theta_0,T}} = L_T^{\theta,\theta_0}$ where*

$$\begin{aligned}
L_T^{\theta,\theta_0} = \exp\Big(&\frac{\alpha - \alpha_0}{4}\log\left(\frac{\det[X_T]}{\det[x]}\right) + \frac{\mathrm{Tr}[bX_T] - \mathrm{Tr}[bx]}{2} - \frac{1}{2}\int_0^T \mathrm{Tr}[b^2 X_s] ds \\
&- \frac{\alpha - \alpha_0}{4}\left(\frac{\alpha + \alpha_0}{2} - 1 - d\right)\int_0^T \mathrm{Tr}[X_s^{-1}] ds - \frac{\alpha T}{2}\mathrm{Tr}[b]\Big).
\end{aligned} \tag{1.7}$$

The proof of Proposition 1.1.1 is given in Appendix 1.7.1, and we assume from now on

$$b \in \mathcal{S}_d.$$

Now, we observe that the quantity in the exponential Equation (1.7) is quadratic with respect to (α, b) and goes almost surely to $-\infty$ when $\|(\alpha, b)\| \to +\infty$. In fact, Cauchy-Schwarz inequality yields to

$$\begin{aligned}
|\mathrm{Tr}[\alpha b]| &= \left|\frac{1}{T}\int_0^T \mathrm{Tr}\left[\sqrt{2} b\sqrt{X_s}\frac{\alpha}{\sqrt{2}}\sqrt{X_s^{-1}}\right] ds\right| \\
&\leqslant \frac{1}{T}\int_0^T \mathrm{Tr}\left[b^2 X_s\right] ds + \frac{\alpha^2}{4}\frac{1}{T}\int_0^T \mathrm{Tr}\left[X_s^{-1}\right] ds,
\end{aligned}$$

and it is strict almost surely, which gives that the quadratic form in the exponential Equation (1.7) is positive definite. There is thus a unique global maximum of Equation (1.7) on $\mathbb{R} \times \mathcal{S}_d$, and the MLE $\hat{\theta}_T = (\hat{\alpha}_T, \hat{b}_T)$ is then characterized by the following equations:

$$\begin{cases} \frac{1}{4}\log\left(\frac{\det[X_T]}{\det[x]}\right) - \frac{\hat{\alpha}_T-1-d}{4}\int_0^T \mathrm{Tr}[X_s^{-1}]ds - \frac{T}{2}\mathrm{Tr}[\hat{b}_T] = 0, \\ \frac{X_T-x}{2} - \frac{1}{2}\int_0^T (\hat{b}_T X_s + X_s \hat{b}_T)ds - \frac{\hat{\alpha}_T T}{2} I_d = 0. \end{cases} \tag{1.8}$$

To get more explicit formulas, we have to invert this linear system. For $X \in \mathcal{S}_d$ and $a \in \mathbb{R}$, we define the linear applications

$$\begin{aligned} \mathcal{L}_X :\ & \mathcal{S}_d \to \mathcal{S}_d & \text{and } \mathcal{L}_{X,a} :\ & \mathcal{S}_d \to \mathcal{S}_d \\ & Y \mapsto YX + XY & & Y \mapsto YX + XY - 2a\,\mathrm{Tr}[Y] I_d. \end{aligned} \tag{1.9}$$

We introduce the following shorthand notation

$$R_T := \int_0^T X_s ds, \quad Q_T := \left(\int_0^T \mathrm{Tr}[X_s^{-1}]ds\right)^{-1}, \quad Z_T := \log\left(\frac{\det[X_T]}{\det[x]}\right), \tag{1.10}$$

and note that Q_T and Z_T are defined only for $\alpha \geqslant d+1$ while R_T is defined for $\alpha \geqslant d-1$ and belongs almost surely to $\mathcal{S}_d^{+,*}$.[1] By using the convexity property of the inverse, see e.g. Mond and Pecaric [56], we have when $\alpha \geqslant d+1$

$$\frac{Q_T^{-1}}{T} < \mathrm{Tr}\left[\left(\frac{R_T}{T}\right)^{-1}\right], \ a.s. \tag{1.11}$$

We get $\hat{\alpha}_T = 1 + d + Q_T(Z_T - 2T\,\mathrm{Tr}[\hat{b}_T])$ and $\mathcal{L}_{R_T, T^2 Q_T}(\hat{b}_T) = X_T - x - T\left(Q_T Z_T + 1 + d\right) I_d$. By Equation (1.11) and Lemma 1.7.1, the latter equation can be inverted, which leads to

$$\begin{cases} \hat{\alpha}_T &= 1 + d + Q_T\Big(Z_T - 2T\,\mathrm{Tr}\left[\mathcal{L}^{-1}_{R_T,T^2Q_T}\left(X_T - x - T\left[Q_T Z_T + 1 + d\right] I_d\right)\right]\Big) \\ \hat{b}_T &= \mathcal{L}^{-1}_{R_T,T^2Q_T}\left(X_T - x - T\left[Q_T Z_T + 1 + d\right] I_d\right). \end{cases} \tag{1.12}$$

The estimator of α when $\alpha \in [d-1, d+1)$ given by the MLE is no longer well defined. The same thing already occurs in dimension $d = 1$ for the CIR process, see Ben Alaya and Kebaier [13]. However, it is still possible to estimate the parameter b when $\alpha \geqslant d-1$ is known. In this case, we denote $\theta = (\alpha, b)$ and $\theta_0 = (\alpha, 0)$ and get by repeating the same arguments that

$$\frac{d\mathbb{P}_{\theta,T}}{d\mathbb{P}_{\theta_0,T}} = \exp\left(\int_0^T \mathrm{Tr}[b\sqrt{X_s}d\tilde{W}_s] - \frac{1}{2}\int_0^T \mathrm{Tr}[b^2 X_s]ds\right).$$

The likelihood and the MLE are then given by

$$L_T^{\theta,\theta_0} = \exp\Big(\frac{\mathrm{Tr}[bX_T] - \mathrm{Tr}[bx]}{2} - \frac{1}{2}\int_0^T \mathrm{Tr}[b^2 X_s]ds - \frac{\alpha T}{2}\mathrm{Tr}[b]\Big), \tag{1.13}$$

$$\hat{b}_T = \mathcal{L}^{-1}_{R_T}\left(X_T - x - \alpha T I_d\right). \tag{1.14}$$

Again, similarly to Proposition 1.1.1, the assumption $b \in \mathcal{S}_d$ is necessary (and sufficient) to have $\frac{d\mathbb{P}_{\theta,T}}{d\mathbb{P}_{\theta_0,T}} \in \mathcal{F}_T^X$.

[1] This is obvious when $\alpha > d-1$ since $X_t \in \mathcal{S}_d^{+,*}$ a.s. by Proposition 4 in [20]. For $\alpha = d-1$, we would have by contradiction the existence of $v_T \in \mathcal{F}_T^X$ such that $\forall t \in [0,T], v_T^\top X_t v_T = 0$. This is clearly not possible by using the connection with matrix-valued Ornstein-Uhlenbeck in this case, see eq. (5.7) in [20].

The goal of the paper is to study the convergence of the MLE under the original probability $\mathbb{P}_\theta$. To do so, we first consider the case where the Wishart process is ergodic, which holds if and only if $-b \in \mathcal{S}_d^{+,*}$ by Lemma 1.7.3. Then, we can use Birkhoff's ergodic theorem to determine the convergence of the MLE. Section 1.2 presents these results for Equation (1.12) when $\alpha \geqslant d+1$ and for Equation (1.14) when $\alpha \geqslant d-1$. Section 1.3 studies the convergence of the MLE in some nonergodic cases, namely when $b = \lambda_0 I_d$ with $\lambda_0 \geqslant 0$. More precisely, when $b = 0$, we obtain convergence results for Equation (1.12) when $\alpha \geqslant d+1$ and for Equation (1.14) when $\alpha \geqslant d-1$. When $\lambda_0 > 0$, we only obtain convergence results for Equation (1.14) when $\alpha \geqslant d-1$. In all these cases, we analyse the convergence by the mean of Laplace transforms. Though limited to some nonergodic cases, we however recover and extend the recent convergence results obtained by Ben Alaya and Kebaier for the one-dimensional CIR process [13,14]. In Section 1.4, we check that the MLE achieves the optimal rate of convergence in the different cases by proving local asymptotic properties. Last, we study in Section 1.5 the Laplace transform of (X_T, R_T). This study can be of independent interest and improves the recent results of Gnoatto and Grasselli [29].

1.2 Statistical Inference of the Wishart process: the ergodic case $-b \in \mathcal{S}_d^{+,*}$

When $-b \in \mathcal{S}_d^{+,*}$, the Wishart process X_t converges in law when $t \to +\infty$ to the stationary law $X_\infty \sim WIS_d(0, \alpha, 0, \sqrt{b^{-1}}; 1/2)$ for any starting point $x \in \mathcal{S}_d^+$ by Lemma 1.7.3. Therefore this is the unique stationary law which is thus extremal, and we know by Stroock ([66], Theorem 7.4.8) that it is then ergodic, see also Pagès [61], Annex A. We introduce the following quantity

$$\overline{R}_\infty := \mathbb{E}_\theta(X_\infty).$$

It is easy to get from Equation (1.4) that $\alpha I_d + b\overline{R}_\infty + \overline{R}_\infty b = 0$, and therefore $\overline{R}_\infty = -\frac{\alpha}{2} b^{-1} \in \mathcal{S}_d^{+,*}$. From the ergodic Birkhoff's theorem, we have

$$\frac{R_T}{T} \xrightarrow{a.s.} \overline{R}_\infty, \qquad \text{as } T \to +\infty. \tag{1.15}$$

Besides, when $\alpha \geqslant d+1$, $\overline{Q}_\infty = \frac{1}{\mathbb{E}_\theta(\mathrm{Tr}[X_\infty^{-1}])}$ is finite and satisfies

$$\overline{Q}_\infty \,\mathrm{Tr}[\overline{R}_\infty^{-1}] < 1, \tag{1.16}$$

due to the convexity property of the inverse, see e.g. Mond and Pecaric [56]. We will show in the proof of Theorem 1.2.1 that

$$\overline{Q}_\infty := \frac{\alpha - (1+d)}{2\,\mathrm{Tr}[-b]}. \tag{1.17}$$

Again, the ergodic Birkhoff's theorem gives

$$TQ_T \xrightarrow{a.s.} \overline{Q}_\infty = \frac{1}{\mathbb{E}_\theta(\mathrm{Tr}[X_\infty^{-1}])}, \qquad \text{as } T \to +\infty. \tag{1.18}$$

1.2.1 The global MLE of $\theta = (\alpha, b)$ when $\alpha \geqslant d+1$

We consider the convergence of the MLE given by Equation (1.12) when $\alpha \geqslant d+1$. We introduce the following martingales:

$$M_t := \int_0^t \sqrt{X_s} dW_s + \int_0^t dW_s^\top \sqrt{X}_s, \tag{1.19}$$

$$N_t := \int_0^t \mathrm{Tr}[(\sqrt{X_s})^{-1} dW_s]. \tag{1.20}$$

We use the dynamics of $(X_t)_{t\geqslant 0}$ under $\mathbb{P}_\theta$ and Itô's formula for $(Z_t)_{t\geqslant 0}$ (see e.g. Bru [20], equation (2.6)) to get on the one hand

$$X_T = x + \alpha T I_d + \mathcal{L}_{R_T}(b) + M_T, \qquad Z_T = (\alpha - 1 - d)Q_T^{-1} + 2\operatorname{Tr}[b]T + 2N_T. \tag{1.21}$$

On the other hand, we obtain from Equation (1.8) and Equation (1.10) that $X_T = x + \hat{\alpha}_T T I_d + \mathcal{L}_{R_T}(\hat{b}_T)$ and $Z_T = (\hat{\alpha}_T - 1 - d)Q_T^{-1} + 2T\operatorname{Tr}[\hat{b}_T]$, which yields to

$$\left\{\begin{array}{rcl} \hat{\alpha}_T - \alpha & = & 2TQ_T\operatorname{Tr}[b - \hat{b}_T] + 2Q_T N_T \\ \\ \mathcal{L}_{R_T}(\hat{b}_T - b) & = & (\alpha - \hat{\alpha}_T)TI_d + M_T = 2T^2 Q_T \operatorname{Tr}[\hat{b}_T - b]I_d + M_T - 2TQ_T N_T. \end{array}\right. \tag{1.22}$$

Theorem 1.2.1. *Assume that $-b \in S_d^{+,*}$ and $\alpha > d+1$. Under $\mathbb{P}_\theta$, $\left(\sqrt{T}(\hat{b}_T - b, \hat{\alpha}_T - \alpha)\right)$ converges in law when $T \to +\infty$ to the centered Gaussian vector $(\mathbf{G}, H)$ that takes values in $\mathcal{S}_d \times \mathbb{R}$ and has the following Laplace transform: for $c, \lambda \in \mathcal{S}_d \times \mathbb{R}$,*

$$\mathbb{E}_\theta\left[\exp\left(\operatorname{Tr}[c\mathbf{G}] + \lambda H\right)\right] = \exp\left(\frac{2\overline{Q}_\infty \lambda^2}{1 - \overline{Q}_\infty \operatorname{Tr}[\overline{R}_\infty^{-1}]} - \frac{2\overline{Q}_\infty \lambda}{1 - \overline{Q}_\infty \operatorname{Tr}[\overline{R}_\infty^{-1}]}\operatorname{Tr}[c\overline{R}_\infty^{-1}] + \operatorname{Tr}[c\mathcal{L}^{-1}_{\overline{R}_\infty, \overline{Q}_\infty}(c)]\right).$$

Proof. By Equation (1.11) and Lemma 1.7.1, we can rewrite the system Equation (1.22) as follows

$$\left\{\begin{array}{rcl} \sqrt{T}(\hat{\alpha}_T - \alpha) & = & 2TQ_T\dfrac{N_T}{\sqrt{T}} - 2TQ_T\operatorname{Tr}\left[\mathcal{L}^{-1}_{\frac{R_T}{T}, TQ_T}\left(\dfrac{M_T}{\sqrt{T}} - 2TQ_T I_d \dfrac{N_T}{\sqrt{T}}\right)\right] \\ \\ \sqrt{T}(\hat{b}_T - b) & = & \mathcal{L}^{-1}_{\frac{R_T}{T}, TQ_T}\left(\dfrac{M_T}{\sqrt{T}} - 2TQ_T I_d \dfrac{N_T}{\sqrt{T}}\right). \end{array}\right.$$

Note that, for $i, j, k, l \in \{1, \ldots, d\}$ we have

$$\begin{array}{rcl} \langle M_{i,j}, M_{k,l}\rangle_t & = & \left[\delta_{jl}(R_t)_{i,k} + \delta_{jk}(R_t)_{i,l} + \delta_{il}(R_t)_{j,k} + \delta_{ik}(R_t)_{j,l}\right], \\ \langle M_{i,j}, N\rangle_t & = & 2t\delta_{ij} \quad \text{and} \quad \langle N\rangle_t = Q_t^{-1}, \end{array} \tag{1.23}$$

where δ_{ij} stands for the Kronecker symbol.
So, it follows from the central limit theorem for martingales (see e.g., Kutoyants [45], Proposition 1.21), that $(\frac{M_T}{\sqrt{T}}, \frac{N_T}{\sqrt{T}})$ converges in law under $\mathbb{P}_\theta$ towards a centered Gaussian vector $(\tilde{\mathbf{G}}, \tilde{H})$ taking values in $\mathcal{S}_d \times \mathbb{R}$ such that

$$\begin{array}{l} \mathbb{E}_\theta(\tilde{\mathbf{G}}_{i,j}\tilde{\mathbf{G}}_{k,l}) = \left[\delta_{jl}(\overline{R}_\infty)_{i,k} + \delta_{jk}(\overline{R}_\infty)_{i,l} + \delta_{il}(\overline{R}_\infty)_{j,k} + \delta_{ik}(\overline{R}_\infty)_{j,l}\right], \\ \mathbb{E}_\theta(\tilde{\mathbf{G}}_{i,j}\tilde{H}) = 2\delta_{i,j} \text{ and } \mathbb{E}_\theta(\tilde{H}^2) = \overline{Q}_\infty^{-1}. \end{array} \tag{1.24}$$

From Equation (1.21) and Equation (1.18), we obtain Equation (1.17). From Lemma 1.7.1, the function $(X, Y, a) \mapsto \mathcal{L}^{-1}_{X,a}(Y)$ is continuous, and we get by Slutsky's theorem that $(\sqrt{T}(\hat{b}_T - b), \sqrt{T}(\hat{\alpha}_T - \alpha))$ converges in law to the Gaussian vector

$$(\mathbf{G}, H) = \left(\mathcal{L}^{-1}_{\overline{R}_\infty, \overline{Q}_\infty}\left(\tilde{\mathbf{G}} - 2\overline{Q}_\infty \tilde{H} I_d\right), 2\overline{Q}_\infty\left(\tilde{H} - \operatorname{Tr}\left[\mathcal{L}^{-1}_{\overline{R}_\infty, \overline{Q}_\infty}\left(\tilde{\mathbf{G}} - 2\overline{Q}_\infty \tilde{H} I_d\right)\right]\right)\right).$$

We are interested to calculate the Laplace transform of this law. First, we calculate the Laplace transform of $(\tilde{\mathbf{G}}, \tilde{H})$:

$$\forall c \in \mathcal{S}_d, \lambda \in \mathbb{R}, \mathbb{E}_\theta\left[\exp\left(\operatorname{Tr}[c\tilde{\mathbf{G}}] + \lambda\tilde{H}\right)\right] = \exp\left(\frac{1}{2}\left(\lambda^2\overline{Q}_\infty^{-1} + 4\lambda\operatorname{Tr}[c] + 4\operatorname{Tr}[c^2\overline{R}_\infty]\right)\right). \tag{1.25}$$

We want to calculate for $c \in \mathcal{S}_d$ and $\lambda \in \mathbb{R}$,

$$\mathbb{E}_\theta\left[\exp\left(\mathrm{Tr}[c\mathbf{G}] + \lambda H\right)\right] = \mathbb{E}_\theta\left[\exp\left(\mathrm{Tr}[(c - 2\lambda\overline{Q}_\infty I_d)\mathbf{G}] + 2\lambda\overline{Q}_\infty\tilde{H}\right)\right].$$

Due to Equation (1.16) and Lemma 1.7.1, we can introduce $\tilde{c} = \mathcal{L}^{-1}_{\overline{R}_\infty,\overline{Q}_\infty}(c - 2\lambda\overline{Q}_\infty I_d)$. We have

$$\overline{R}_\infty\tilde{c} + \tilde{c}\overline{R}_\infty - 2\overline{Q}_\infty\,\mathrm{Tr}[\tilde{c}]I_d = c - 2\lambda\overline{Q}_\infty I_d,$$

and thus

$$\begin{aligned}\mathrm{Tr}[(c - 2\lambda\overline{Q}_\infty I_d)\mathbf{G}] &= \mathrm{Tr}[(\overline{R}_\infty\tilde{c} + \tilde{c}\overline{R}_\infty - 2\overline{Q}_\infty\,\mathrm{Tr}[\tilde{c}]I_d)\mathbf{G}]\\ &= \mathrm{Tr}[\tilde{c}(\overline{R}_\infty\mathbf{G} + \mathbf{G}\overline{R}_\infty - 2\overline{Q}_\infty\,\mathrm{Tr}[\mathbf{G}]I_d)] = \mathrm{Tr}[\tilde{c}(\tilde{\mathbf{G}} - 2\overline{Q}_\infty\tilde{H}I_d)].\end{aligned}$$

We therefore obtain from Equation (1.25)

$$\begin{aligned}&\mathbb{E}_\theta\left[\exp\left(\mathrm{Tr}[c\mathbf{G}] + \lambda H\right)\right]\\ &= \mathbb{E}_\theta\left[\exp\left(\mathrm{Tr}[\tilde{c}(\tilde{\mathbf{G}} - 2\overline{Q}_\infty\tilde{H}I_d)] + 2\lambda\overline{Q}_\infty\tilde{H}\right)\right]\\ &= \mathbb{E}_\theta\left[\exp\left(\mathrm{Tr}[\tilde{c}\tilde{\mathbf{G}}] + 2\overline{Q}_\infty(\lambda - \mathrm{Tr}[\tilde{c}])\tilde{H}\right)\right]\\ &= \exp\left(2\left\{(\lambda - \mathrm{Tr}[\tilde{c}])^2\overline{Q}_\infty + 2(\lambda - \mathrm{Tr}[\tilde{c}])\,\mathrm{Tr}[\tilde{c}]\overline{Q}_\infty + \mathrm{Tr}[\tilde{c}^2\overline{R}_\infty]\right\}\right).\end{aligned}$$

Since $2\,\mathrm{Tr}[\tilde{c}^2\overline{R}_\infty] = \mathrm{Tr}[\tilde{c}(\tilde{c}\overline{R}_\infty + \overline{R}_\infty\tilde{c})] = \mathrm{Tr}[\tilde{c}c] + 2\overline{Q}_\infty(\mathrm{Tr}[\tilde{c}] - \lambda)\,\mathrm{Tr}[\tilde{c}]$, we get

$$\mathbb{E}_\theta\left[\exp\left(\mathrm{Tr}[c\mathbf{G}] + \lambda H\right)\right] = \exp\left(2\lambda(\lambda - \mathrm{Tr}[\tilde{c}])\overline{Q}_\infty + \mathrm{Tr}[\tilde{c}c]\right).$$

We now use that $\mathcal{L}^{-1}_{\overline{R}_\infty,\overline{Q}_\infty}(I_d) = \frac{1}{2(1-\overline{Q}_\infty\,\mathrm{Tr}[\overline{R}_\infty^{-1}])}\overline{R}_\infty^{-1}$ to get $\tilde{c} = \mathcal{L}^{-1}_{\overline{R}_\infty,\overline{Q}_\infty}(c) - \lambda\frac{\overline{Q}_\infty\overline{R}_\infty^{-1}}{1-\overline{Q}_\infty\,\mathrm{Tr}[\overline{R}_\infty^{-1}]}$. Since we have $\mathrm{Tr}[\mathcal{L}^{-1}_{\overline{R}_\infty,\overline{Q}_\infty}(c)] = \frac{\mathrm{Tr}[\overline{R}_\infty^{-1}c]}{2(1-\overline{Q}_\infty\,\mathrm{Tr}[\overline{R}_\infty^{-1}])}$ by Lemma 1.7.1, this yields to the claimed result. $\square$

When $\alpha = d+1$, the rate of convergence of the MLE of α is even better as stated by the following theorem.

Theorem 1.2.2. *Assume $-b \in S_d^{+,*}$ and $\alpha = d+1$. Then, under $\mathbb{P}_\theta$, $\left(\sqrt{T}(\hat{b}_T - b), T(\hat{\alpha}_T - \alpha)\right)$ converges in law when $T \to +\infty$ to $\left(\mathbf{G}, -2\tau^{-1}_{-\mathrm{Tr}[b]}\,\mathrm{Tr}[b + \mathbf{G}]\right)$, where $\tau_a = \inf\{t \geq 0,\ B_t = a\}$ with $(B_t)_{t\geq 0}$ a given one-dimensional standard Brownian motion and $\mathbf{G}$ is a Gaussian vector independent of B such that $\mathbb{E}_\theta\left[\exp\left(\mathrm{Tr}[c\mathbf{G}]\right)\right] = \exp\left(\mathrm{Tr}[c\mathcal{L}^{-1}_{\overline{R}_\infty}(c)]\right)$, $c \in \mathcal{S}_d$.*

Proof. By Equation (1.11) and Lemma 1.7.1, we can rewrite the system Equation (1.22) as follows

$$\left\{\begin{array}{rcl} T(\hat{\alpha}_T - \alpha) &=& 2T^2Q_T\left(\dfrac{N_T}{T} - \mathrm{Tr}\left[\mathcal{L}^{-1}_{\frac{R_T}{T},TQ_T}\left(\dfrac{M_T}{\sqrt{T}} - 2T^{3/2}Q_TI_d\dfrac{N_T}{T}\right)\right]\right)\\[2ex] \sqrt{T}(\hat{b}_T - b) &=& \mathcal{L}^{-1}_{\frac{R_T}{T},TQ_T}\left(\dfrac{M_T}{\sqrt{T}} - 2T^{3/2}Q_TI_d\dfrac{N_T}{T}\right).\end{array}\right. \tag{1.26}$$

From Equation (1.21), we have

$$\frac{N_T}{T} = \frac{1}{2T}\log\left(\frac{\det[X_T]}{\det[x]}\right) - \mathrm{Tr}[b].$$

As for $-b \in S_d^{+,*}$ the Wishart process $(X_t)_{t\geq 0}$ is stationary with invariant limit distribution X_∞ we easily deduce that $\frac{N_T}{T}$ converges in probability to $-\operatorname{Tr}[b]$ when $T \to \infty$. Then, it follows from Equation (1.15) that

$$(T^{-1}R_T, T^{-1}N_T) \overset{\mathbb{P}_\theta}{\to} (\overline{R}_\infty, -\operatorname{Tr}[b]), \quad as\ T \to \infty. \tag{1.27}$$

Hence, we only need to study the asymptotic behavior of the couple $(T^{-1/2}M_T, T^2 Q_T)$. According to Theorem 4.1 in Mayerhofer [54], we have for $\lambda \geqslant 0$ and $\Gamma \in S_d$

$$\mathbb{E}_\theta\left[\exp\left(\frac{\lambda}{T}N_T - \frac{\lambda^2}{2T^2}Q_T^{-1} + \frac{1}{\sqrt{T}}\operatorname{Tr}[\Gamma M_T] - \frac{2}{T}\int_0^T \operatorname{Tr}[\Gamma^2 X_s]ds - \frac{2\lambda}{\sqrt{T}}\operatorname{Tr}[\Gamma]\right)\right] = 1. \tag{1.28}$$

Now, let us introduce the quantity

$$\begin{aligned} A_T = \mathbb{E}_\theta\Bigg[\exp\left(\lambda\frac{N_T}{T} + \lambda\operatorname{Tr}[b]\right)\exp\left(-\frac{\lambda^2}{2T^2}Q_T^{-1} + \frac{1}{\sqrt{T}}\operatorname{Tr}[\Gamma M_T]\right) \\ \times \exp\left(-\frac{2}{T}\int_0^T \operatorname{Tr}[\Gamma^2 X_s]ds + 2\operatorname{Tr}[\Gamma^2\overline{R}_\infty]\right)\Bigg]. \end{aligned}$$

Then, by Equation (1.28) we easily get $A_T = \exp\left(\lambda\operatorname{Tr}[b] + 2\operatorname{Tr}[\Gamma^2\overline{R}_\infty] + \frac{2\lambda}{\sqrt{T}}\operatorname{Tr}[\Gamma]\right)$. We now write $A_T = \tilde{A}_T + \mathbb{E}_\theta\left[\exp\left(-\frac{\lambda^2}{2T^2}Q_T^{-1} + \frac{1}{\sqrt{T}}\operatorname{Tr}[\Gamma M_T]\right)\right]$ with

$$\begin{aligned} \tilde{A}_T &= \mathbb{E}_\theta\left[(\exp(\xi_T) - 1)\exp\left(-\frac{\lambda^2}{2T^2}Q_T^{-1} + \frac{1}{\sqrt{T}}\operatorname{Tr}[\Gamma M_T]\right)\right] \\ \xi_T &= \lambda\frac{N_T}{T} + \lambda\operatorname{Tr}[b] - \frac{2}{T}\int_0^T \operatorname{Tr}[\Gamma^2 X_s]ds + 2\operatorname{Tr}[\Gamma^2\overline{R}_\infty]. \end{aligned}$$

Cauchy-Schwarz inequality and $Q_T^{-1} > 0$ give

$$|\tilde{A}_T| \leqslant \mathbb{E}_\theta^{1/2}[\exp(2\xi_T) - 2\exp(\xi_T) + 1]\mathbb{E}_\theta^{1/2}\left[\exp\left(\frac{2}{\sqrt{T}}\operatorname{Tr}[\Gamma M_T]\right)\right].$$

On the one hand, Proposition 1.5.1 with $m = -b \in \mathcal{S}_d^{+,*}$ gives

$$\mathbb{E}_\theta\left[\exp\left(\frac{2}{\sqrt{T}}\operatorname{Tr}[\Gamma M_T]\right)\right] \leqslant \mathbb{E}_\theta\left[\exp\left(\frac{2}{T}\operatorname{Tr}[\Gamma^2 R_T]\right)\right] < \infty.$$

On the other hand, we have for any $r \geqslant 0$,

$$\mathbb{E}_\theta[\exp(r\xi_T)] \leqslant \mathbb{E}_\theta\left[\exp\left(\frac{\lambda r}{T}N_T\right)\right]\exp(2r\operatorname{Tr}[\Gamma^2\overline{R}_\infty]).$$

From Equation (1.21), we have

$$\mathbb{E}_\theta\left[\exp\left(\frac{\lambda r}{T}N_T\right)\right] = \exp(-\lambda r\operatorname{Tr}[b])\mathbb{E}_\theta\left[\left(\frac{\det[X_T]}{\det[x]}\right)^{\frac{\lambda r}{2T}}\right].$$

The sublinear growth of the coefficients of the Wishart SDE and the convergence to a stationary law gives that $\mathbb{E}_\theta\left[\left(\frac{\det[X_T]}{\det[x]}\right)^{\tilde{\lambda}}\right]$ is uniformly bounded in $T > 0$, $\tilde{\lambda} < 1$ and therefore $\sup_{T > \frac{\lambda r}{2}}\mathbb{E}_\theta\left[\left(\frac{\det[X_T]}{\det[x]}\right)^{\frac{\lambda r}{2T}}\right] < \infty$. This gives the uniform integrability of the family $(\exp(2\xi_T), T >$

λ). Then, we deduce from Equation (1.27) that $\mathbb{E}_\theta[\exp(2\xi_T) - 2\exp(\xi_T) + 1] \underset{T\to+\infty}{\to} 0$ and thus $\tilde{A}_T \underset{T\to+\infty}{\to} 0$.
Hence, we obtain

$$\lim_{T\to\infty} \mathbb{E}_\theta \left[\exp\left(-\frac{\lambda^2}{2T^2} Q_T^{-1} + \frac{1}{\sqrt{T}} \operatorname{Tr}[\Gamma M_T] \right) \right] = \lim_{T\to\infty} A_T = \exp\left(\lambda \operatorname{Tr}[b] + 2\operatorname{Tr}[\Gamma^2 \overline{R}_\infty] \right).$$

Therefore, we deduce by Lemma 1.7.2 the following convergence in law

$$\left(\frac{Q_T^{-1}}{T^2}, \frac{M_T}{\sqrt{T}} \right) \Rightarrow \left(\tau_{-\operatorname{Tr}[b]}, \sqrt{\overline{R}_\infty}\tilde{\mathbf{G}} + \tilde{\mathbf{G}}^\top \sqrt{\overline{R}_\infty} \right) \text{ as } T \to \infty,$$

where $\tilde{G}_{i,j}$ $1 \leqslant i,j \leqslant d$ are independent standard normal variables. Together with Equation (1.27), we obtain that

$$(T^{-1}R_T, T^2 Q_T, T^{-1}N_T, T^{-1/2}M_T) \Rightarrow (\overline{R}_\infty, 1/\tau_{-\operatorname{Tr}[b]}, -\operatorname{Tr}[b], \sqrt{\overline{R}_\infty}\tilde{\mathbf{G}} + \tilde{\mathbf{G}}^\top \sqrt{\overline{R}_\infty}), \tag{1.29}$$

which gives the claim by Equation (1.26) and Lemma 1.7.2. $\square$

1.2.2 The MLE estimator of b when $\alpha \geqslant d11$

When $\alpha \in [d-1, d+1)$, we are no longer able to study the convergence of the MLE of α. It is however still possible to get the speed of convergence of the MLE of b.

Theorem 1.2.3. *Assume that $-b \in S_d^{+,*}$ and $\alpha \geqslant d-1$. For $T > 0$, we consider $\hat{b}_T$ defined by (1.14). Then, under $\mathbb{P}_\theta$, $\sqrt{T}(\hat{b}_T - b)$ converges in law to a centered Gaussian vector $\mathbf{G}$ with the following Laplace transform $\mathbb{E}_\theta[\exp(\operatorname{Tr}[c\mathbf{G}])] = \exp\left(\operatorname{Tr}[c\mathcal{L}_{\overline{R}_\infty}^{-1}(c)]\right)$, $c \in \mathcal{S}_d$.*

Proof. We could prove this result by using the explicit Laplace transform Proposition 1.5.1. Here, we use the same arguments as before based on the ergodic property. From Equation (1.14), we have

$$\sqrt{T}(\hat{b}_T - b) = \mathcal{L}_{\frac{R_T}{T}}^{-1}\left(\frac{1}{\sqrt{T}}(X_T - x - bR_T - R_T b - \alpha T I_d) \right) = \mathcal{L}_{\frac{R_T}{T}}^{-1}\left(\frac{M_T}{\sqrt{T}} \right).$$

As in the proof of Theorem 1.2.1, $\frac{M_T}{\sqrt{T}}$ converges in law to the centered Gaussian vector $\tilde{\mathbf{G}}$ defined by Equation (1.24). Slutsky's theorem and Equation (1.15) give then the convergence of $\sqrt{T}(\hat{b}_T - b)$ to $\mathbf{G} = \mathcal{L}_{\overline{R}_\infty}^{-1}\left(\tilde{\mathbf{G}}\right)$, whose Laplace transform is given by Lemma 1.7.2. $\square$

1.3 Statistical Inference of the Wishart process: some nonergodic cases

This section studies the convergence of the MLE in the case $b = b_0 I_d$ with $b_0 \geqslant 0$. When $b_0 = 0$ and $\alpha \geqslant d+1$, we are able to describe the rate of convergence of the MLE of (α, b) given by Equation (1.12). When $b_0 > 0$ and $\alpha \geqslant d-1$, we can also obtain the rate of convergence of the MLE of b given by Equation (1.14). Last, when b is known a priori to be diagonal, the MLE of b has a simpler form and we can describe precisely its convergence.

1.3.1 The global MLE of $\theta=(\alpha,b)$ when $b=0$ and $\alpha\geqslant d+1$

The following result provides the asymptotic behavior of the estimator of the couple when $\alpha>d+1$ and $b=0$ in (1.4).

Theorem 1.3.1. *Assume that $b=0$ and $\alpha>d+1$. Let $(\hat b_T,\hat\alpha_T)$ be the MLE defined by (1.12). Then, $(T(\hat b_T-b),\sqrt{\log(T)}(\hat\alpha_T-\alpha))$ converges in law under $\mathbb{P}_\theta$ when $T\to+\infty$ to*

$$\left(\mathcal{L}^{-1}_{R^0_1}\left(X^0_1-\alpha I_d\right),2\sqrt{\frac{\alpha-(d+1)}{d}}G\right),$$

where $X^0_t=\alpha t I_d+\int_0^t\sqrt{X^0_s}dW_s+dW_s^\top\sqrt{X^0_s}$ is a Wishart process with the same parameters but starting from 0, $R^0_t=\int_0^t X^0_s ds$ and $G\sim\mathcal{N}(0,1)$ is an independent standard Normal variable.

Proof. From Equation (1.12) and Equation (1.22), we obtain

$$\begin{cases}\sqrt{\log(T)}(\hat\alpha_T-\alpha) &= -2\frac{T\operatorname{Tr}[\hat b_T]}{\sqrt{\log(T)}}\log(T)Q_T+2\log(T)Q_T\frac{N_T}{\sqrt{\log(T)}}\\ T\hat b_T &= \mathcal{L}^{-1}_{\frac{R_T}{T^2},Q_T}\left(\frac{X_T}{T}-\frac{x}{T}-(Q_TZ_T+1+d)I_d\right),\end{cases}$$

and we are interested in studying the convergence in law of $\left(\frac{N_T}{\sqrt{\log(T)}},\frac{X_T}{T},\frac{R_T}{T^2}\right)$. By Theorem 4.1 in [54], for $\mu\geqslant 0$ and $T>1$,

$$\frac{d\overline{\mathbb{P}}}{d\mathbb{P}_\theta}=\exp\left(\frac{\mu N_T}{\sqrt{\log(T)}}-\frac{\mu^2}{2Q_T\log(T)}\right)$$

defines a change of probability and $(X_t)_{t\in[0,T]}$ is a Wishart process with degree $\alpha+\frac{\mu}{\sqrt{\log(T)}}$ under $\overline{\mathbb{P}}$. Let $\lambda_1,\lambda_2\in\mathcal{S}_d^{+,*}$ and

$$A_T=\mathbb{E}_\theta\left[\exp\left(\frac{\mu N_T}{\sqrt{\log(T)}}-\frac{\mu^2}{2Q_T\log(T)}\right)\exp\left(-\operatorname{Tr}\left[\frac{\lambda_2}{T}X_T\right]-\operatorname{Tr}\left[\frac{\lambda_1}{T^2}R_T\right]\right)\right].$$

By Proposition 1.5.1, we have

$$\begin{aligned}A_T&=\overline{\mathbb{E}}\left[\exp\left(-\operatorname{Tr}\left[\frac{\lambda_2}{T}X_T\right]-\operatorname{Tr}\left[\frac{\lambda_1}{T^2}R_T\right]\right)\right]\\&=\frac{1}{\det[V]^{\frac{\alpha+\mu/\sqrt{\log(T)}}{2}}}\exp\left(-\frac{1}{2T}\operatorname{Tr}\left[V'V^{-1}x\right]\right)\underset{T\to+\infty}{\to}\frac{1}{\det[V]^{\frac{\alpha}{2}}},\end{aligned}\tag{1.30}$$

where

$$V=(\sqrt{2\lambda_1})^{-1}\sinh(\sqrt{2\lambda_1})2\lambda_2+\cosh(\sqrt{2\lambda_1}),\ V'=2\cosh(\sqrt{2\lambda_1})\lambda_2+\sqrt{2\lambda_1}\sinh(\sqrt{2\lambda_1}).\tag{1.31}$$

We note that this limit does not depend on μ and is the Laplace transform of (X^0_1,R^0_1) by Proposition 1.5.1.
We now use that $\frac{1}{Q_T\log(T)}\underset{T\to+\infty}{\to}\frac{d}{\alpha-(d+1)}$ a.s., see Lemma 1.7.4 and we define

$$\tilde A_T=\mathbb{E}_\theta\left[\exp(\xi_T)\right]\ \text{and}\ \xi_T=\frac{\mu N_T}{\sqrt{\log(T)}}-\frac{\mu^2 d}{2(\alpha-(d+1))}-\operatorname{Tr}\left[\frac{\lambda_2}{T}X_T\right]-\operatorname{Tr}\left[\frac{\lambda_1}{T^2}R_T\right]$$

that is finite by using equation Equation (1.62) of Lemma 1.7.4 since $\xi_T \leqslant \frac{\mu N_T}{\sqrt{\log(T)}}$. We have

$$A_T = \tilde{A}_T + \mathbb{E}_\theta \left[\left\{ \exp\left(\frac{\mu^2 d}{2(\alpha - (d+1))} - \frac{\mu^2}{2Q_T \log(T)} \right) - 1 \right\} \exp(\xi_T) \right].$$

The Cauchy-Schwarz inequality gives

$$(A_T - \tilde{A}_T)^2 \leqslant \mathbb{E}_\theta \left[\left\{ \exp\left(\frac{\mu^2 d}{2(\alpha - (d+1))} - \frac{\mu^2}{2Q_T \log(T)} \right) - 1 \right\}^2 \right] \mathbb{E}_\theta \left[\exp(2\xi_T) \right].$$

Since $Q_T \log(T)$ is positive for $T > 1$ and converges a.s. to $\frac{\alpha-(d+1)}{d}$, the first expectation goes to 0 while the second one is bounded by using again Equation (1.62). Therefore, $A_T - \tilde{A}_T \underset{T\to+\infty}{\to} 0$, and we get

$$\mathbb{E}_\theta \left[\exp\left(\frac{\mu N_T}{\sqrt{\log(T)}} - \mathrm{Tr}\left[\frac{\lambda_2}{T} X_T\right] - \mathrm{Tr}\left[\frac{\lambda_1}{T^2} R_T\right] \right) \right] \underset{T\to+\infty}{\to} \frac{\exp\left(\frac{\mu^2 d}{2(\alpha-(d+1))}\right)}{\det[V]^{\frac{\alpha}{2}}}.$$

Thus, $\left(\frac{N_T}{\sqrt{\log(T)}}, \frac{X_T}{T}, \frac{R_T}{T^2} \right)$ converges in law to $(\sqrt{\frac{d}{\alpha-(d+1)}} G, X_1^0, R_1^0)$, where $G \sim \mathcal{N}(0,1)$ is independent of X^0. From Equation (1.21), we have

$$Q_T Z_T + 1 + d = 2\frac{1}{\sqrt{\log(T)}} \log(T) Q_T \frac{N_T}{\sqrt{\log(T)}} + \alpha,$$

and therefore $Q_T Z_T + 1 + d$ converges in probability to α. Slutsky's theorem gives then the following convergence in law: as $T \to +\infty$,

$$\left(\frac{N_T}{\sqrt{\log(T)}}, \frac{X_T}{T}, \frac{R_T}{T^2}, Q_T Z_T + 1 + d, Q_T \log(T) \right) \Rightarrow \left(\sqrt{\frac{d}{\alpha - (d+1)}} G, X_1^0, R_1^0, \alpha, \frac{\alpha - (d+1)}{d} \right). \tag{1.32}$$

This gives the claimed convergence for $(\hat{\alpha}_T, \hat{b}_T)$ due to the continuity property given in Lemma 1.7.1.

□

Theorem 1.3.2. *Assume that $b = 0$ and $\alpha = d+1$. Let $(\hat{b}_T, \hat{\alpha}_T)$ be the MLE defined by (1.12). Then, $(T(\hat{b}_T - b), \log(T)(\hat{\alpha}_T - \alpha))$ converges in law under $\mathbb{P}_\theta$ when $T \to +\infty$ to*

$$\left(\mathcal{L}_{R_1^0}^{-1}\left(X_1^0 - \alpha I_d \right), \frac{4}{d\tau_1} \right),$$

where $X_t^0 = \alpha t I_d + \int_0^t \sqrt{X_s^0} dW_s + dW_s^\top \sqrt{X_s^0}$ is a Wishart process with the same parameters but starting from 0, $R_t^0 = \int_0^t X_s^0 ds$ and $\tau_1 = \inf\{t \geqslant 0, B_t = 1\}$ where B is a standard Brownian motion independent from W.

Proof. The proof follows the same line as the one of Theorem 1.3.1, but we now write

$$\log(T)(\hat{\alpha}_T - \alpha) = -2\frac{T\,\mathrm{Tr}[\hat{b}_T]}{\log(T)} \log(T)^2 Q_T + 2\log(T)^2 Q_T \frac{N_T}{\log(T)},$$

while we still have $T\hat{b}_T = \mathcal{L}^{-1}_{\frac{R_T}{T^2}, Q_T}\left(\frac{X_T}{T} - \frac{x}{T} - (Q_T Z_T + 1 + d) I_d \right)$. By Theorem 4.1 in [54], for $\mu \geqslant 0$ and $T > 1$, $\frac{d\overline{\mathbb{P}}}{d\mathbb{P}} = \exp\left(\frac{\mu N_T}{\log(T)} - \frac{\mu^2}{2Q_T \log(T)^2} \right)$ defines a change of probability, and we define for $\lambda_1, \lambda_2 \in \mathcal{S}_d^{+,*}$,

$$A_T = \mathbb{E}_\theta \left[\exp\left(\frac{\mu N_T}{\log(T)} - \frac{\mu^2}{2Q_T \log(T)^2} \right) \exp\left(-\mathrm{Tr}\left[\frac{\lambda_2}{T} X_T\right] - \mathrm{Tr}\left[\frac{\lambda_1}{T^2} R_T\right] \right) \right].$$

By Proposition 1.5.1, we have

$$A_T = \frac{1}{\det[V]^{\frac{\alpha+\mu/\log(T)}{2}}} \exp\Big(-\frac{1}{2T}\operatorname{Tr}\left[V'V^{-1}x\right]\Big) \underset{T\to+\infty}{\to} \frac{1}{\det[V]^{\frac{\alpha}{2}}},$$

where V and V' are defined by Equation (1.31).
We now use that $\frac{N_T}{\log(T)} \to \frac{d}{2}$ in probability, see Lemma 1.7.4, and define

$$\tilde{A}_T = \mathbb{E}_\theta\left[\exp(\xi_T)\right],\ \xi_T = \frac{\mu d}{2} - \frac{\mu^2}{2Q_T\log(T)^2} - \operatorname{Tr}\left[\frac{\lambda_2}{T}X_T\right] - \operatorname{Tr}\left[\frac{\lambda_1}{T^2}R_T\right]$$

and have

$$A_T = \tilde{A}_T + \mathbb{E}_\theta\left[\left\{\exp\left(\frac{\mu N_T}{\log(T)} - \frac{\mu d}{2}\right) - 1\right\}\exp\left(\xi_T\right)\right].$$

We note that $\exp(\xi_T) \leqslant \exp\left(\frac{\mu d}{2}\right)$. By using Lemma 1.7.4 and the uniform integrability Equation (1.63), we get that $A_T - \tilde{A}_T \underset{T\to+\infty}{\to} 0$ and therefore

$$\mathbb{E}_\theta\left[\exp\left(-\frac{\mu^2}{2Q_T\log(T)^2} - \operatorname{Tr}\left[\frac{\lambda_2}{T}X_T\right] - \operatorname{Tr}\left[\frac{\lambda_1}{T^2}R_T\right]\right)\right] \underset{T\to+\infty}{\to} \frac{\exp(-\mu d/2)}{\det[V]^{\frac{\alpha}{2}}}.$$

Therefore, $\left(\frac{X_T}{T}, \frac{R_T}{T^2}, Q_T\log(T)^2\right)$ converges in law to $\left(X_1^0, R_1^0, \left(\frac{2}{d}\right)^2\frac{1}{\tau_1}\right)$, where τ_1 is independent of X^0. We observe that $Q_TZ_T = \frac{1}{\log(T)}Q_T\log(T)^2\frac{Z_T}{\log(T)}$. Lemma 1.7.4 and Slutsky's theorem gives

$$\left(\frac{N_T}{\log(T)}, \frac{X_T}{T}, \frac{R_T}{T^2}, Q_TZ_T + 1 + d, Q_T\log(T)^2\right) \Rightarrow \left(\frac{d}{2}, X_1^0, R_1^0, d+1, \left(\frac{2}{d}\right)^2\frac{1}{\tau_1}\right), \tag{1.33}$$

which gives the claim by Lemma 1.7.1. □

1.3.2 The MLE of b when $\alpha \geqslant d-1$

Until the end of this section we consider that $\alpha \geqslant d-1$ is known and study the speed of convergence of the estimator of b defined by Equation (1.14).

Case $b = 0$.

Theorem 1.3.3. *Assume that $b = 0$ and $\alpha \geqslant d-1$. For $T > 0$, let $\hat{b}_T$ be defined by (1.14). When $T \to +\infty$, $T(\hat{b}_T - b)$ converges in law under $\mathbb{P}_\theta$ to $\mathcal{L}^{-1}_{R_1^0}\left(X_1^0 - \alpha I_d\right)$, where $(X_t^0)_{t\geqslant 0}$ is the solution to $X_t^0 = \alpha t I_d + \int_0^t \sqrt{X_s^0}dW_s + dW_s^\top\sqrt{X_s^0}$ and $R_t^0 = \int_0^t X_s^0 ds$.*

Proof. From Equation (1.14), we have $T\hat{b}_T = \mathcal{L}^{-1}_{\frac{R_T}{T^2}}\left(\frac{X_T}{T} - \frac{x}{T} - \alpha I_d\right)$. Let V and V' be defined by Equation (1.31). Similarly to Equation (1.30), we have by Proposition 1.5.1 for $\lambda_1, \lambda_2 \in \mathcal{S}_d^{+,*}$

$$\mathbb{E}_\theta\left[\exp\left(-\operatorname{Tr}\left[\frac{\lambda_2}{T}X_T\right] - \operatorname{Tr}\left[\frac{\lambda_1}{T^2}R_T\right]\right)\right] = \frac{1}{\det[V]^{\frac{\alpha}{2}}}\exp\Big(-\frac{1}{2T}\operatorname{Tr}\left[V'V^{-1}x\right]\Big) \underset{T\to+\infty}{\to} \frac{1}{\det[V]^{\frac{\alpha}{2}}}.$$

This gives the convergence in law of $\left(\frac{X_T}{T}, \frac{R_T}{T^2}\right)$ to (X_1^0, R_1^0) and then the claimed result. □

Case $b = b_0 I_d, b_0 > 0$.

In this case $b = b_0 I_d$ with $b_0 > 0$. In order to identify the speed of convergence and the limit law, we use the Laplace transform approach. We have the following result,

Theorem 1.3.4. *Assume that $b = b_0 I_d$, $b_0 > 0$, and $\alpha \geqslant d-1$. For $T > 0$ let $\hat{b}_T$ defined by (1.14). When $T \to +\infty$, $\exp(b_0 T)(\hat{b}_T - b)$ converges in law under $\mathbb{P}_\theta$ to $\mathcal{L}_X^{-1}\left(\sqrt{X}\tilde{\mathbf{G}} + \tilde{\mathbf{G}}\sqrt{X}\right)$ where $X \sim WIS_d\left(\frac{x}{2b_0}, \alpha, 0, I_d; \frac{1}{4b_0^2}\right)$ and $\tilde{\mathbf{G}}$ is an independent d-square matrix whose elements are independent standard Normal variables.*

The proof of this results relies on the explicit calculation of the Laplace transform of (X_T, R_T) and is postponed to Subsection 1.5.2.
Obviously, the case $b = b_0 I_d$ is very particular. One would like to consider more general nonergodic cases or ideally to be able to state a general convergence results of $\hat{b}_T$ towards b for any $b \in \mathcal{S}_d$. Despite our efforts, we have not been able to get such a result. The reason why we can handle the ergodic case and the nonergodic case $b = b_0 I_d$ with $b_0 \geqslant 0$ is that the convergence of all the matrix terms occurs at the same speed, namely $1/\sqrt{T}$ for the ergodic case, $1/T$ for $b = 0$ and $e^{-b_0 T}$ when $b_0 > 0$. In the other cases, there is no such a simple scalar rescaling. Heuristically, there may be different speeds of convergence that are difficult to disentangle because of the different matrix products. To get an idea of this, we present now the case of the estimation of b when b is known to be a diagonal matrix. In this case, we obtain different speed of convergence for each diagonal terms.

The MLE of b when b is known a priori to be diagonal.

We assume that $\alpha \geqslant d-1$ is known and that b is a diagonal matrix, i.e. $b = \mathrm{diag}(b_1, \cdots, b_d)$. We want to estimate the diagonal elements by maximizing the likelihood. We denote $\theta_0 = (\alpha, 0)$. As in Equation (1.13), we have

$$L_T^{\theta,\theta_0} = \exp\Big(\frac{\mathrm{Tr}[bX_T] - \mathrm{Tr}[bx]}{2} - \frac{1}{2}\int_0^T \mathrm{Tr}[b^2 X_s]ds - \frac{\alpha T}{2}\mathrm{Tr}[b]\Big).$$

By differentiating this with respect to b_i, $1 \leqslant i \leqslant d$, we get

$$\frac{\partial_{b_i} L_T^{\theta,\theta_0}}{L_T^{\theta,\theta_0}} = \frac{1}{2}\Big((X_T)_{i,i} - x_{i,i} - \alpha t I_d - 2b_i \int_0^T (X_s)_{i,i} ds\Big),$$

and therefore the MLE of b is given by

$$(\hat{b}_T)_i = \frac{(X_T)_{i,i} - x_{i,i} - \alpha T}{2(R_T)_{i,i}}. \tag{1.34}$$

We therefore obtain

$$(\hat{b}_T)_i - b_i = \frac{(X_T)_{i,i} - x_{i,i} - \alpha T - 2b_i (R_T)_{i,i}}{2(R_T)_{i,i}}. \tag{1.35}$$

Let us observe that this estimator is precisely the one obtained by Ben Alaya and Kebaier [14] for the CIR process. This is not very surprising since we know from Equation (1.4), Equation (1.2) and b diagonal that there exists independent Brownian motions β^i, $1 \leqslant i \leqslant d$ such that

$$d(X_t)_{i,i} = (\alpha + 2b_i (X_t)_{i,i})dt + 2\sqrt{(X_t)_{i,i}}d\beta_t^i.$$

Thus, the diagonal elements follow independent CIR processes, and the observation of the non diagonal elements does not improve the ML estimation. We can obtain the asymptotic convergence by applying Theorem 1 in [13], up to a small correction in the nonergodic case which is given by our Theorem 1.3.4 in dimension $d = 1$. This yields to the following proposition.

Proposition 1.3.1. *Let $\alpha \geqslant d-1$ and b a diagonal matrix. Let $\epsilon_t = \operatorname{diag}(\epsilon_t^1, \cdots, \epsilon_t^d)$ be a diagonal matrix with*

$$\epsilon_t^i = \begin{cases} t^{-\frac{1}{2}} & \text{if} \quad b_i < 0 \\ t^{-1} & \text{if} \quad b_i = 0 \\ \exp(-b_i t) & \text{if} \quad b_i > 0 \end{cases}$$

Then, under $\mathbb{P}_\theta$, $\epsilon_T^{-1} \operatorname{diag}((\hat{b}_T)_1 - b_1, \cdots, (\hat{b}_T)_d - b_d)$ converges in law to a diagonal matrix $\mathbf{D}$ made with independent elements. Each diagonal element $\mathbf{D}_i$ is distributed as follows:

$$\forall i \in \{1, \cdots, d\}, \quad \mathbf{D}_i \underset{law}{=} \begin{cases} \sqrt{\frac{-2b_i}{\alpha}}\mathbf{G} & \text{if } b_i < 0 \\ \frac{X_1^0 - \alpha}{2R_1^0} & \text{if } b_i = 0 \\ \frac{\mathbf{G}}{\sqrt{X_{1/(4b_i^2)}^{x_{i,i}/(2b_i)}}}, & \text{if } b_i > 0 \end{cases}$$

where $X_t^x = x + \alpha t + 2\int_0^t \sqrt{X_s^x} dW_s$, $R_t^0 = \int_0^t X_s^0 ds$, and $\mathbf{G} \sim \mathcal{N}(0,1)$ is independent of X.

1.4 Optimality of the MLE

In parametric estimation theory, a fundamental role is played by the local asymptotic normality (**LAN**) property since the work of Le Cam [46]. This general concept developed by Le Cam is extended later by Le Cam and Yang [47] and Jeganathan [37] to local asymptotic mixed normality (**LAMN**) and local asymptotic quadraticity (**LAQ**) properties. These notions are mainly dedicated to study the asymptotic efficiency of estimators of a given parametric model. The aim of this section is to check the validity of either **LAN**, **LAMN** or **LAQ** properties for the global model in order to get the asymptotic efficiency of our maximum likelihood estimators studied in the previous section. Here we prove these properties only for the global model $\theta = (\alpha, b)$. The same technique applies for the submodel related to the estimation of b and for each treated case we have been able to obtain the corresponding local asymptotic property.
Let us consider the Wishart process $(X_t)_{t \geqslant 0} \in \mathcal{S}_d^+$ with parameters $\theta := (\alpha, b)$, with $\alpha \geq d+1$ and $b \in \mathcal{S}_d$.

$$\begin{cases} dX_t = [\alpha I + bX_t + X_t b]\, dt + \sqrt{X}_t dW_t + dW_t^\top \sqrt{X}_t, t > 0 \\ X_0 \in \mathcal{S}_d^+. \end{cases} \tag{1.36}$$

We recall that $\mathbb{P}_\theta$ denotes the distributions induced by the solutions of Equation (1.36) on canonical space $C(\mathbb{R}_+, \mathcal{S}_d^+)$ with the natural filtration $\mathcal{F}_t^X := \sigma(X_s, s \leq t)$ and $\mathbb{P}_{\theta,t} = \mathbb{P}_{\theta|\mathcal{F}_t^X}$ denotes the restriction of $\mathbb{P}_\theta$ on the filtration $\mathcal{F}_t$.
For $\tilde{\alpha} \geq d+1$ and $\tilde{b} \in \mathcal{S}_d^+$, we set $\tilde{\theta} = (\tilde{\alpha}, \tilde{b})$,

$$H_t = \frac{\tilde{\alpha} - \alpha}{2} (\sqrt{X_t})^{-1} + (\tilde{b} - b)\sqrt{X_t},$$

and we introduce the log-likelihood function

$$\ell_T^\theta(\tilde{\theta}) \;=\; \log\left(\frac{d\mathbb{P}_{\tilde{\theta},T}}{d\mathbb{P}_{\theta,T}}\right) = \int_0^T \operatorname{Tr}[H_s dW_s] - \frac{1}{2}\int_0^T \operatorname{Tr}[H_s H_s^\top] ds \tag{1.37}$$

The process $(\tilde{W}_t = W_t - \int_0^t H_s^\top ds, t \leq T)$ is a $d \times d$-Brownian motion under $\mathbb{P}_{\tilde{\theta},T}$. In the sequel, let us introduce the quantity $\delta_T := (\delta_{1,T}, \delta_{2,T}) \in \mathbb{R}^2$ where for $i \in \{1,2\}$ the localizing rates satisfy

$|\delta_{i,T}| \to 0$ when $T \to \infty$. For all $u := (u_1, u_2) \in \mathbb{R} \times \mathcal{S}_d$, we define $\delta_T \cdot u := (\delta_{1,T} u_1, \delta_{2,T} u_2) \in \mathbb{R} \times \mathcal{S}_d$. Now, we rewrite Equation (1.37) with $\tilde{\theta} = \theta + \delta_T \cdot u$

$$\begin{aligned}\ell_T^{\theta}(\theta + \delta_T \cdot u) &= \int_0^T \operatorname{Tr}\left[\frac{\delta_{1,T} u_1 (\sqrt{X_s})^{-1}}{2} dW_s + \delta_{2,T} u_2 \sqrt{X_s} dW_s\right] \\ &\quad -\frac{1}{2}\int_0^T \operatorname{Tr}\left[\delta_{2,T} u_2 X_s \delta_{2,T} u_2\right] ds - \frac{T}{2}\delta_{1,T} u_1 \operatorname{Tr}\left[\delta_{2,T} u_2\right] \\ &\quad -\frac{1}{8}\int_0^T (\delta_{1,T} u_1)^2 \operatorname{Tr}\left[(X_s)^{-1}\right] ds.\end{aligned}$$

Hence, by using the definitions Equation (1.10), Equation (1.19) and Equation (1.20) of the martingales processes $(N_t)_{t\geq 0}$ and $(M_t)_{t\geq 0}$ and the processes $(R_t)_{t\geq 0}$ and $(Q_t)_{t\geq 0}$, it is easy to check that

$$\begin{aligned}\ell_T^{\theta}(\theta + \delta_T \cdot u) &= \frac{1}{2}\Big(\delta_{1,T} u_1 N_T + \operatorname{Tr}\big[\delta_{2,T} u_2 M_T\big]\Big) - \frac{1}{2}\operatorname{Tr}\left[\delta_{2,T} u_2 R_T \delta_{2,T} u_2\right] \\ &\quad -\frac{T}{2}\delta_{1,T} u_1 \operatorname{Tr}\left[\delta_{2,T} u_2\right] - \frac{1}{8}(\delta_{1,T} u_1)^2 (Q_T)^{-1} \\ &= \ell_T^{\theta}(\theta + \delta_T \cdot u) = \Lambda_T(u) - \frac{1}{2}\Gamma_T(u), \end{aligned} \tag{1.38}$$

where $\Lambda_T(u) := \frac{1}{2}\Big(\delta_{1,T} u_1 N_T + \operatorname{Tr}\big[\delta_{2,T} u_2 M_T\big]\Big)$ is a linear random function with respect to $u \in \mathbb{R} \times \mathcal{S}_d^+$ with quadratic variation

$$\Gamma_T(u) := \delta_{2,T}^2 \operatorname{Tr}[u_2^2 R_T] + T\delta_{1,T}\delta_{2,T} u_1 \operatorname{Tr}\left[u_2\right] + \frac{1}{4}\delta_{1,T}^2 u_1^2 Q_T^{-1}.$$

1.4.1 Case $-b \in S_d^+$ and $\alpha \geqslant d+1$

We first consider $\alpha > d+1$. In this ergodic case, we set $\delta_{i,T} = T^{-1/2}$ for $i \in \{1, 2\}$, and we get from Equation (1.15) and Equation (1.18)

$$\Gamma_T(u) \xrightarrow{a.s} \overline{\Gamma}_\infty(u) := \operatorname{Tr}\left[u_2^2 \overline{R}_\infty\right] + u_1 \operatorname{Tr}\left[u_2\right] + \frac{1}{4}u_1^2 \operatorname{Tr}\left[(\overline{Q}_\infty)^{-1}\right], \quad \text{as } T \to +\infty. \tag{1.39}$$

This yields the validity of the so called Raykov type condition. Hence, according to Theorem 1 in [52], relations Equation (1.38) and Equation (1.39) ensure the validity of the local asymptotic normality (**LAN**) property, that is under $\mathbb{P}_\theta$ we have

$$(\Lambda_T(u), \Gamma_T(u)) \Rightarrow \left(\overline{\Gamma}_\infty^{1/2}(u) Z, \overline{\Gamma}_\infty(u)\right), \text{ as } T \to \infty, \tag{1.40}$$

with Z a standard normal real random variable. It is worth noting that the above convergence can also be obtained using the proof of Theorem 1.2.1. In fact, we have already proven that under $\mathbb{P}_\theta$

$$\left(\frac{N_T}{\sqrt{T}}, \frac{M_T}{\sqrt{T}}\right) \Rightarrow (\tilde{\mathbf{G}}, \tilde{H}) \tag{1.41}$$

where $(\tilde{G}, \tilde{H})$ is a centered Gaussian vector taking values in $\mathbb{R} \times S_d^+$ such that

$$\begin{aligned}\mathbb{E}_\theta(\tilde{\mathbf{G}}_{i,j}\tilde{\mathbf{G}}_{k,l}) &= \left[\delta_{jl}(\overline{R}_\infty)_{i,k} + \delta_{jk}(\overline{R}_\infty)_{i,l} + \delta_{il}(\overline{R}_\infty)_{j,k} + \delta_{ik}(\overline{R}_\infty)_{j,l}\right], \\ \mathbb{E}_\theta(\tilde{\mathbf{G}}_{i,j}\tilde{H}) &= 2\delta_{i,j} \text{ and } \mathbb{E}_\theta(\tilde{H}^2) = \overline{Q}_\infty^{-1}.\end{aligned}$$

Therefore, **LAN** property Equation (1.40) follows from relations Equation (1.39) and Equation (1.41).

We now consider the case $\alpha = d+1$ and set $\delta_{1,T} = T^{-1}$ and $\delta_{2,T} = T^{-1/2}$. By using Equation (1.29), we get that under $\mathbb{P}_\theta$,

$$(\Lambda_T(u), \Gamma_T(u)) \Rightarrow \left(-\frac{u_1}{2} \operatorname{Tr}[b] + \operatorname{Tr}\left[u_2 \sqrt{\overline{R}_\infty} \tilde{\mathbf{G}} \right], \operatorname{Tr}[u_2^2 \overline{R}_\infty] + \frac{1}{4} u_1^2 \tau_{-\operatorname{Tr}[b]} \right), \quad \text{as } T \to \infty,$$

where $\tau_{-\operatorname{Tr}[b]}$ is defined as in Theorem 1.2.2 and $\tilde{G}$ is an independent matrix, whose elements $\tilde{G}_{i,j}$, $1 \leqslant i, j \leqslant d$, are independent standard normal variables. Hence, according to Le Cam and Yang [47] and Jeganathan [37] this last convergence yields the **LAQ** property for this ergodic case.

1.4.2 Case $b = 0$ and $\alpha \geqslant d + 1$

We first assume $\alpha > d + 1$. From Equation (1.36) with $b = 0$ and Equation (1.19), we have $M_T = X_T - x - \alpha I_d T$. From Equation (1.32), it follows that as $T \to \infty$

$$\left(\frac{N_T}{\sqrt{\log(T)}}, \frac{M_T}{T}, \frac{R_T}{T^2}, \frac{Q_T^{-1}}{\log(T)} \right) \Rightarrow \left(\sqrt{\frac{d}{\alpha - (d+1)}} G, X_1^0 - \alpha I_d, R_1^0, \frac{d}{\alpha - (d+1)} \right),$$

where X_1^0 and R_1^0 are defined as in Theorem 1.3.1. Thus, in the same way as in the previous case if we set $\delta_{1,T} = \frac{1}{\sqrt{\log(T)}}$ and $\delta_{2,T} = T^{-1}$, then $(\Lambda_T(u), \Gamma_T(u))$ converges in law under $\mathbb{P}_\theta$ to

$$\left(\frac{1}{2} \sqrt{\frac{d}{\alpha - (d+1)}} u_1 G + \frac{1}{2} \operatorname{Tr}\left[u_2 (X_1^0 - \alpha I_d) \right], \operatorname{Tr}\left[u_2^2 \overline{R}_1^0 \right] + \frac{u_1^2 d}{4(\alpha - (d+1))} \right), \quad \text{as } T \to \infty.$$

This ensures the validity of the **LAQ** property in this non-ergodic case.
When $\alpha = d + 1$, we use the notation of Theorem 1.3.2 and get from Equation (1.33)

$$\left(\frac{N_T}{\log(T)}, \frac{M_T}{T}, \frac{R_T}{T^2}, \frac{Q_T^{-1}}{\log(T)^2} \right) \Rightarrow \left(\frac{d}{2}, X_1^0 - \alpha I_d, R_1^0, \left(\frac{d}{2} \right)^2 \tau_1 \right).$$

With $\delta_{1,T} = \frac{1}{\log(T)}$ and $\delta_{2,T} = T^{-1}$, we get that $(\Lambda_T(u), \Gamma_T(u))$ converges in law under $\mathbb{P}_\theta$ to

$$\left(\frac{d}{4} u_1 + \frac{1}{2} \operatorname{Tr}[u_2 (X_1^0 - \alpha I_d)], \operatorname{Tr}\left[u_2^2 \overline{R}_1^0 \right] + \frac{d^2 u_1^2}{8} \tau_1 \right), \quad \text{as } T \to \infty.$$

This gives again the **LAQ** property.

1.5 The Laplace transform and its use to study the MLE

1.5.1 The Laplace transform of (X_T, R_T)

We present our main result on the joint Laplace transform of (X_T, R_T), that can be of independent interest. This Laplace transform is given by Bru [20], eq. (4.7) when $b = 0$ and has been recently studied and obtained explicitly by Gnoatto and Grasselli [29]. Here, we present another proof that enables us to get the Laplace transform for any $\alpha \geqslant d - 1$, as well as a more precise result concerning its set of convergence, see Remarks 1.5.1 and 1.5.2 below for a further discussion.

Proposition 1.5.1. *Let $\alpha \geqslant d-1$, $x \in \mathcal{S}_d^+$, $b \in \mathcal{S}_d$ and $X \sim WIS_d(x,\alpha,b,I_d)$. Let $v,w \in \mathcal{S}_d$ be such that*

$$\exists m \in \mathcal{S}_d,\ \frac{v}{2} - mb - bm - 2m^2 \in \mathcal{S}_d^+ \text{ and } \frac{w}{2} + m \in \mathcal{S}_d^+. \tag{1.42}$$

Then, we have for $t \geqslant 0$

$$\begin{aligned}&\mathbb{E}\Big[\exp\Big(-\frac{1}{2}\operatorname{Tr}[wX_t] - \frac{1}{2}\operatorname{Tr}[vR_t]\Big)\Big]\\ &= \frac{\exp\big(-\frac{\alpha}{2}\operatorname{Tr}[b]t\big)}{\det[V_{v,w}(t)]^{\frac{\alpha}{2}}}\exp\Big(-\frac{1}{2}\operatorname{Tr}\left[(V'_{v,w}(t)V_{v,w}(t)^{-1}+b)x\right]\Big),\end{aligned} \tag{1.43}$$

with

$$V_{v,w}(t) = \left(\sum_{k=0}^{\infty} t^{2k+1}\frac{\tilde{v}^k}{(2k+1)!}\right)\tilde{w} + \sum_{k=0}^{\infty} t^{2k}\frac{\tilde{v}^k}{(2k)!}, \quad \tilde{v} = v + b^2, \quad \text{and} \quad \tilde{w} = w - b.$$

If besides $\tilde{v} = v + b^2 \in \mathcal{S}_d^{+,*}$, we have $V_{v,w}(t) = (\sqrt{\tilde{v}})^{-1}\sinh(\sqrt{\tilde{v}}t)\tilde{w} + \cosh(\sqrt{\tilde{v}}t)$ and then $V'_{v,w}(t) = \cosh(\sqrt{\tilde{v}}t)\tilde{w} + \sinh(\sqrt{\tilde{v}}t)\sqrt{\tilde{v}}$.

Before proving this result, we recall the following fact:

$$\forall x,y \in \mathcal{S}_d^+,\ \operatorname{Tr}[xy] \geqslant 0, \tag{1.44}$$

which is clear once we have observed that $\operatorname{Tr}[xy] = \operatorname{Tr}[\sqrt{x}y\sqrt{x}]$ and $\sqrt{x}y\sqrt{x} \in \mathcal{S}_d^+$. We also recall a result on matrix Riccati equations, see Dieci and Eirola [24] Proposition 1.1.

Lemma 1.5.1. *Let $\tilde{b} \in \mathcal{S}_d$ and $\tilde{\delta} \in \mathcal{S}_d^+$. Let ξ denote the solution of the following matrix Riccati differential equation*

$$\xi' + 2\xi^2 = \tilde{b}\xi + \xi\tilde{b} + \tilde{\delta},\ \xi(0) \in \mathcal{S}_d. \tag{1.45}$$

If $\xi(0) \in \mathcal{S}_d^+$, the solution ξ is well-defined for any $t \geqslant 0$ and satisfies $\xi(t) \in \mathcal{S}_d^+$.

Proof of Proposition 1.5.1. Let $T > 0$ be given. We first assume $w,v \in \mathcal{S}_d^+$, which ensures that $\mathbb{E}\left[\exp\left(-\frac{1}{2}\operatorname{Tr}[wX_T + vR_T]\right)\right] < \infty$. We consider the martingale

$$M_t = \mathbb{E}\left[\exp\left(-\frac{1}{2}\operatorname{Tr}[wX_T + vR_T]\right)\Big|\mathcal{F}_t\right],\ t \in [0,T].$$

Due to the affine structure, we are looking for smooth functions $\beta : \mathbb{R}_+ \to \mathbb{R}$, $\gamma,\delta : \mathbb{R}_+ \to \mathcal{S}_d$ such that

$$M_t = \exp\left(\beta(T-t) + \operatorname{Tr}[\gamma(T-t)X_t] + \operatorname{Tr}[\delta(T-t)R_t]\right).$$

We necessarily have $\beta(0) = 0$, $\gamma(0) = -w/2$ and $\delta(0) = -v/2$. Itô's formula gives

$$\begin{aligned}\frac{dM_t}{M_t} =& \Big\{-\beta'(T-t) - \operatorname{Tr}[\gamma'(T-t)X_t] - \operatorname{Tr}[\delta'(T-t)R_t] + \operatorname{Tr}[\gamma(T-t)(\alpha I_d + bX_t + X_tb)]\\ &+ \operatorname{Tr}[\delta(T-t)X_t] + 2\operatorname{Tr}[\gamma(T-t)^2X_t]\Big\}dt + \operatorname{Tr}[\gamma(T-t)(\sqrt{X_t}dW_t + dW_t^\top\sqrt{X_t})].\end{aligned}$$

Since M is a martingale, the drift term should vanish almost surely, almost everywhere. The drift term being a (deterministic) affine function of (X_t,R_t), we obtain the following system of differential equations:

$$\delta' = 0, \tag{1.46}$$

$$-\gamma' + \gamma b + b\gamma + 2\gamma^2 + \delta = 0, \tag{1.47}$$

$$-\beta' + \alpha\operatorname{Tr}[\gamma] = 0. \tag{1.48}$$

The first equation gives $\delta(t) = -v/2$. The second equation is a matrix Riccati differential equation. We now consider $\xi = m - \gamma$ with m satisfying Equation (1.42). It solves Equation (1.45) with $\tilde{b} = b + 2m$, $\tilde{\delta} = -bm - mb - 2m^2 + v/2$ and $\xi(0) = m + w/2$. We know then by Lemma 1.5.1 that ξ is well defined for any $t \geqslant 0$ and stays in $\mathcal{S}_d^+$. In particular, γ is well defined for any $t \geqslant 0$. We set $\tilde{\gamma} = \gamma + \frac{1}{2}b$. We have $\gamma^2 = \tilde{\gamma}^2 - \frac{1}{2}(b\tilde{\gamma} + \tilde{\gamma}b) + \frac{1}{4}b^2$ and thus $\tilde{\gamma}$ solves the following matrix Riccati differential equation:

$$\tilde{\gamma}' = 2\tilde{\gamma}^2 - \frac{1}{2}\tilde{v},\ \tilde{\gamma}(0) = -\frac{1}{2}\tilde{w},\ \text{with } \tilde{v} = v + b^2 \text{ and } \tilde{w} = w - b.$$

We set $M(t) = \begin{bmatrix} M_1(t) & M_2(t) \\ M_3(t) & M_4(t) \end{bmatrix} = \exp\left(t \begin{bmatrix} 0 & -\tilde{v}/2 \\ -2I_d & 0 \end{bmatrix}\right) \in \mathcal{M}_{2d}$ and get by Levin [49] that

$$\tilde{\gamma}(t) = \left[M_2(t) - \frac{1}{2}M_1(t)\tilde{w}\right]\left[M_4(t) - \frac{1}{2}M_3(t)\tilde{w}\right]^{-1}.$$

We check that the matrix $M_4(t) - \frac{1}{2}M_3(t)\tilde{w}$ is indeed invertible. In fact, let

$$\tau = \inf\left\{t \geqslant 0, \det\left[M_4(t) - \frac{1}{2}M_3(t)\tilde{w}\right] = 0\right\}.$$

We have $\tau > 0$ and for $t \in [0, \tau)$, $\frac{d}{dt}[M_4(t) - \frac{1}{2}M_3(t)\tilde{w}] = -2\left[M_2(t) - \frac{1}{2}M_1(t)\tilde{w}\right]$ and thus

$$\frac{d}{dt}\det\left[M_4(t) - \frac{1}{2}M_3(t)\tilde{w}\right] = -2\det\left[M_4(t) - \frac{1}{2}M_3(t)\tilde{w}\right]\mathrm{Tr}[\tilde{\gamma}(t)].$$

This gives $\det\left[M_4(t) - \frac{1}{2}M_3(t)\tilde{w}\right] = \exp(-2\int_0^t \mathrm{Tr}[\tilde{\gamma}(s)]ds) > 0$, and we necessary get $\tau = +\infty$ since γ and thus $\tilde{\gamma}$ is well defined for $t \geqslant 0$.
Since

$$\begin{bmatrix} 0 & -\tilde{v}/2 \\ -2I_d & 0 \end{bmatrix}^{2k} = \begin{bmatrix} \tilde{v}^k & 0 \\ 0 & \tilde{v}^k \end{bmatrix} \text{ and } \begin{bmatrix} 0 & -\tilde{v}/2 \\ -2I_d & 0 \end{bmatrix}^{2k+1} = \begin{bmatrix} 0 & -\tilde{v}^{k+1}/2 \\ -2\tilde{v}^k & 0 \end{bmatrix},$$

we get

$$M_1(t) = M_4(t) = \sum_{k=0}^{\infty} \frac{t^{2k}\tilde{v}^k}{(2k)!},\ M_2(t) = -\frac{1}{2}\sum_{k=0}^{\infty} \frac{t^{2k+1}\tilde{v}^{k+1}}{(2k+1)!},\ M_3(t) = -2\sum_{k=0}^{\infty} \frac{t^{2k+1}\tilde{v}^k}{(2k+1)!}.$$

If $\tilde{v} = v + b^2 \in \mathcal{S}_d^{+,*}$, $\sqrt{\tilde{v}}$ is well defined and we have $M_1(t) = M_4(t) = \cosh(t\sqrt{\tilde{v}})$, $M_2(t) = -\frac{1}{2}\sqrt{\tilde{v}}\sinh(t\sqrt{\tilde{v}})$ and $M_3(t) = -2(\sqrt{\tilde{v}})^{-1}\sinh(t\sqrt{\tilde{v}})$. Now, we define

$$V(t) = M_4(t) - \frac{1}{2}M_3(t)\tilde{w} = \left(\sum_{k=0}^{\infty} t^{2k+1}\frac{\tilde{v}^k}{(2k+1)!}\right)\tilde{w} + \sum_{k=0}^{\infty} t^{2k}\frac{\tilde{v}^k}{(2k)!}.$$

Since $V'(t) = -2M_2(t) + M_1(t)\tilde{w}$, we obtain that

$$\tilde{\gamma}(t) = -\frac{1}{2}V'(t)V(t)^{-1} \text{ and thus } \gamma(t) = -\frac{1}{2}\left(V'(t)V(t)^{-1} + b\right).$$

Last, we have $\beta'(t) = -\frac{1}{2}\alpha\,\mathrm{Tr}[V'(t)V(t)^{-1}] - \frac{1}{2}\alpha\,\mathrm{Tr}[b]$ and we obtain that

$$\beta(t) = -\frac{1}{2}\alpha\log(\det[V(t)]) - \frac{1}{2}\alpha\,\mathrm{Tr}[b]t,$$

since $\frac{d\det[V(t)]}{dt} = \det[V(t)]\operatorname{Tr}[V'(t)V(t)^{-1}]$.
It remains to show that we indeed have Equation (1.43) for v and w satisfying Equation (1.42). We define $\mathcal{E}_t = \frac{\exp\left(\beta(T-t)+\operatorname{Tr}[\gamma(T-t)X_t]+\operatorname{Tr}[-\frac{v}{2}R_t]\right)}{\exp(\beta(T)+\operatorname{Tr}[\gamma(T)x])}$. By Itô's formula, we have

$$\frac{d\mathcal{E}_t}{\mathcal{E}_t} = \operatorname{Tr}[\gamma(T-t)(\sqrt{X_t}dW_t + dW_t^\top\sqrt{X_t})].$$

This is a positive local martingale and thus a supermartingale which gives $\mathbb{E}[\mathcal{E}_T] \leqslant 1$, and we want to prove that this is a martingale. To do so, we use the argument presented by Rydberg in [65]. For $L > 0$, we define

$$\tau_L = \inf\{t \geqslant 0, \operatorname{Tr}[X_t] \geqslant L\},$$

and $\pi_L(x) = x\mathbf{1}_{\operatorname{Tr}[x]\leqslant L} + \frac{L}{\operatorname{Tr}[x]}x\mathbf{1}_{\operatorname{Tr}[x]>L}$ for $x \in \mathcal{S}_d^+$. We consider $(\mathcal{E}_t^L, t \in [0,T])$ the solution of

$$d\mathcal{E}_t^L = \mathcal{E}_t^L\operatorname{Tr}[\gamma(T-t)(\sqrt{\pi_L(X_t)}dW_t + dW_t^\top\sqrt{\pi_L(X_t)})],\ \mathcal{E}_0^L = 1.$$

We clearly have $\mathbb{E}[\mathcal{E}_T^L] = 1$. Besides, under $\mathbb{P}^L$ given by $\frac{d\mathbb{P}^L}{d\mathbb{P}}|_{\mathcal{F}_T} = \mathcal{E}_T^L$, the process

$$dW_t^L = dW_t - 2\sqrt{\pi_L(X_t)}\gamma(T-t)dt,\ t \in [0,T],$$

is a matrix Brownian motion. Since $\mathcal{E}_t = \mathcal{E}_t^L$ for $t \leqslant \tau_L$, we have $\mathbb{E}[\mathcal{E}_T] = \mathbb{E}[\mathcal{E}_T^L\mathbf{1}_{\tau_L>T}] + \mathbb{E}[\mathcal{E}_T\mathbf{1}_{\tau_L\leqslant T}]$. By Lebesgue's theorem, we get $\mathbb{E}[\mathcal{E}_T\mathbf{1}_{\tau_L\leqslant T}] \underset{L\to+\infty}{\to} 0$. On the other hand, $\mathbb{E}[\mathcal{E}_T^L\mathbf{1}_{\tau_L>T}] = \mathbb{P}^L(\tau_L > T)$. Let us consider the Wishart process $\tilde{X}$ starting from x such that

$$d\tilde{X}_t = \Big[\alpha I_d + (b+2\gamma(T-t))\tilde{X}_t + \tilde{X}_t(b+2\gamma(T-t))\Big]dt + \sqrt{\tilde{X}_t}dW_t + dW_t^\top\sqrt{\tilde{X}_t}.$$

We also define $\tilde{\tau}_L = \inf\{t \in [0,T], \operatorname{Tr}[\tilde{X}_t] \geqslant L\}$ with convention $\inf\emptyset = +\infty$. The process $\tilde{X}$ solves the same SDE on $[0,\tilde{\tau}_L\wedge T]$ under $\mathbb{P}$ as X on $[0,\tau_L\wedge T]$ under $\mathbb{P}^L$. We therefore have

$$\mathbb{P}^L(\tau_L > T) = \mathbb{P}(\tilde{\tau_L} > T) \underset{L\to+\infty}{\to} 1,$$

which finally gives $\mathbb{E}[\mathcal{E}_T] = 1$. $\square$

Corollary 1.5.1. *Let $Y \sim WIS_d(y,\alpha,b,a)$ be a Wishart process with parameters $\alpha \geqslant d-1$, $y \in \mathcal{S}_d^+$, $a, b \in \mathcal{M}_d$ satisfying*

$$ba^\top a = a^\top ab^\top \text{ and } a \text{ invertible.} \tag{1.49}$$

Let $v, w \in \mathcal{S}_d$ be such that

$$\exists m \in \mathcal{S}_d,\ \frac{1}{2}awa^\top + m \in \mathcal{S}_d^+ \text{ and } \frac{ava^\top}{2} - ab^\top a^{-1}m - m(a^\top)^{-1}ba^\top - 2m^2 \in \mathcal{S}_d^+. \tag{1.50}$$

Then, we have

$$\begin{aligned}&\mathbb{E}\left[\exp\left(-\frac{1}{2}\operatorname{Tr}\left[wY_T + v\int_0^T Y_s ds\right]\right)\right]\\ &= \frac{\exp\left(-\frac{\alpha}{2}\operatorname{Tr}[b]t\right)}{\det[V_{v,w}(t)]^{\frac{\alpha}{2}}}\exp\Big(-\frac{1}{2}\operatorname{Tr}\left[(V_{v,w}'(t)V_{v,w}(t)^{-1} + (a^\top)^{-1}ba^\top)(a^\top)^{-1}ya^{-1}\right]\Big),\end{aligned}$$

with $V_{v,w}(t) = \left(\sum_{k=0}^\infty t^{2k+1}\frac{\tilde{v}^k}{(2k)!}\right)\tilde{w} + \sum_{k=0}^\infty t^{2k}\frac{\tilde{v}^k}{(2k)!}$ and

$$\tilde{v} = ava^\top + (a^\top)^{-1}b^2a^\top, \quad \text{and} \quad \tilde{w} = awa^\top - (a^\top)^{-1}ba^\top.$$

Proof. We know that $Y \underset{law}{=} a^\top X a$ with $x = (a^\top)^{-1} y a^{-1}$ and $X \sim WIS_d(x, \alpha, (a^\top)^{-1} b a^\top, I_d)$, see e.g. equation (13) in [1]. We notice that $(a^\top)^{-1} b a^\top = a b^\top a^{-1} \iff b a^\top a = a^\top a b^\top$ and thus $(a^\top)^{-1} b a^\top \in \mathcal{S}_d$. We have

$$\mathbb{E}\left[\exp\left(-\frac{1}{2}\operatorname{Tr}\left[wY_T + v\int_0^T Y_s ds\right]\right)\right] = \mathbb{E}\left[\exp\left(-\frac{1}{2}\operatorname{Tr}\left[awa^\top X_T + ava^\top \int_0^T X_s ds\right]\right)\right],$$

which gives the result by applying Proposition 1.5.1. $\square$

By setting $\tilde{m} = a^{-1} m (a^\top)^{-1}$, the condition Equation (1.50) is equivalent to the existence of $\tilde{m} \in \mathcal{S}_d$, such that

$$\frac{1}{2}w + \tilde{m} \in \mathcal{S}_d^+ \text{ and } \frac{v}{2} - b^\top \tilde{m} - \tilde{m} b - 2\tilde{m} a^\top a \tilde{m} \in \mathcal{S}_d^+. \tag{1.51}$$

The case $m = 0$ gives back the finiteness of the Laplace transform when $v, w \in \mathcal{S}_d^+$. If we take $\tilde{m} = -w/2$, we get also the finiteness when

$$v + b^\top w + wb - wa^\top a w \in \mathcal{S}_d^+. \tag{1.52}$$

Another interesting choice is $m = -\frac{1}{2}(a^\top)^{-1} b a^\top$. We have $m \in \mathcal{S}_d$ from Equation (1.49). This choice gives the finiteness of the Laplace transform when $v + b^\top (a^\top a)^{-1} b \in \mathcal{S}_d^+$ and $w - (a^\top a)^{-1} b \in \mathcal{S}_d^+$. Let us note that $\tilde{v} = a(v + b^\top (a^\top a)^{-1} b) a^\top$ so that the first condition is the same as $\tilde{v} \in \mathcal{S}_d^+$. Another interesting choice of m is given by the next remark.

Remark 1.5.1. *Proposition 1.5.1 extends the result of Gnoatto and Grasselli [29] to $\alpha \geqslant d-1$, and the sufficient condition Equation (1.50) that ensures the finiteness of the Laplace transform is also less restrictive, which is crucial in our study especially in the nonergodic case. In particular, it does not assume a priori that $v + b^\top (a^\top a)^{-1} b \in \mathcal{S}_d^+$. We can recover the result of [29] as follows. Let us assume $v + b^\top (a^\top a)^{-1} b \in \mathcal{S}_d^+$ and take $m = -\frac{(a^\top)^{-1} b a^\top}{2} + \frac{1}{2}\sqrt{a(v + b^\top (a^\top a)^{-1} b) a^\top}$. We have $m \in \mathcal{S}_d$ from Equation (1.49) and it satisfies $\frac{ava^\top}{2} - ab^\top a^{-1} m - m (a^\top)^{-1} b a^\top - 2m^2 = 0 \in \mathcal{S}_d^+$. Therefore, Equation (1.50) holds if*

$$w - (a^\top a)^{-1} b + a^{-1}\sqrt{a(v + b^\top (a^\top a)^{-1} b) a^\top}\,(a^\top)^{-1} \in \mathcal{S}_d^+.$$

This is precisely the condition stated in [29].

Remark 1.5.2. *It is possible to get similarly the Laplace transform of $\left(Y_T, \int_0^T Y_s ds\right)$ when Y solves*

$$dY_t = \left[\overline{\alpha} + bY_t + Y_t b^\top\right] dt + \sqrt{Y}_t dW_t a + a^\top dW_t^\top \sqrt{Y}_t,\ Y_0 = y \in \mathcal{S}_d^+,$$

with a, b satisfying Equation (1.49) and $\overline{\alpha} - (d-1) a^\top a \in \mathcal{S}_d^+$. Again, equation (13) in [1] gives $Y \underset{law}{=} a^\top X a$, where

$$dX_t = \left[\hat{\alpha} + \hat{b} X_t + X_t \hat{b}^\top\right] dt + \sqrt{X}_t dW_t + dW_t^\top \sqrt{X}_t,\ X_0 = x,$$

with $x = (a^\top)^{-1} y a^{-1} \in \mathcal{S}_d$, $\hat{b} = (a^\top)^{-1} b a^\top \in \mathcal{S}_d$ and $\hat{\alpha} = (a^\top)^{-1} \overline{\alpha} a^{-1} \in \mathcal{S}_d$. Repeating the proof of Proposition 1.5.1, we observe that the Riccati equation Equation (1.47) and equation Equation (1.46) remain unchanged while Equation (1.48) is replaced by

$$\beta' = \operatorname{Tr}[\hat{\alpha}\gamma] = -\frac{1}{2}\operatorname{Tr}[\hat{\alpha} V'(t) V(t)^{-1}] - \frac{1}{2}\operatorname{Tr}[\hat{\alpha}\hat{b}].$$

Therefore, we deduce that under the same condition Equation (1.50), we have

$$\begin{aligned}&\mathbb{E}\left[\exp\left(-\frac{1}{2}\operatorname{Tr}\left[wY_T + v\int_0^T Y_s ds\right]\right)\right]\\&= \exp(\beta(T))\exp\Big(-\frac{1}{2}\operatorname{Tr}\left[(V'_{v,w}(t)V_{v,w}(t)^{-1} + (a^\top)^{-1}ba^\top)(a^\top)^{-1}ya^{-1}\right]\Big),\end{aligned}$$

with $\beta(t) = -\frac{1}{2}\int_0^t \operatorname{Tr}[(a^\top)^{-1}\overline{\alpha}a^{-1}V'_{v,w}(s)V_{v,w}(s)^{-1}]ds - \frac{t}{2}\operatorname{Tr}[\overline{\alpha}(a^\top a)^{-1}b]$ *and* $V_{v,w}(t)$ *defined as in Corollary 1.5.1. Thus, the formula is no longer totally explicit. In Gnoatto and Grasselli [29], the result is stated with* $\operatorname{Tr}[(a^\top)^{-1}\overline{\alpha}a^{-1}\log(V_{v,w}(t))]$ *instead of the first integral. However, this replacement does not seem clear to us unless* $V'_{v,w}(s)$ *and* $V_{v,w}(s)$ *commute for all* $s \geqslant 0$ *(this happens when the matrices* $\tilde{v}$ *and* $\tilde{w}$ *in* $V_{v,w}$ *commute) or* $\overline{\alpha} = \alpha a^\top a$ *by using the trace cyclic theorem.*

Corollary 1.5.2. *Let* $Y \sim WIS_d(y, \alpha, b, a)$ *be a Wishart process with parameters such that* $ba^\top a = a^\top ab^\top$ *and* a *invertible. Then,*

$$\forall u \in \mathcal{S}_d,\ \mathbb{E}\left[\exp\left(\int_0^T \operatorname{Tr}[u\sqrt{Y_s}dW_s a]ds - \frac{1}{2}\int_0^T \operatorname{Tr}[auY_s ua^\top]ds\right)\right] = 1.$$

Proof. We have $2\int_0^T \operatorname{Tr}[u\sqrt{Y_s}dW_s a]ds = \operatorname{Tr}[u(Y_T - y)] - \alpha T\operatorname{Tr}[ua^\top a] - \operatorname{Tr}\left[(ub + b^\top u)\int_0^T Y_s ds\right]$. We apply Corollary 1.5.1 with $w = -u$ and $v = ub + b^\top u + ua^\top au$. Therefore, Equation (1.52) holds. We then have $\tilde{w} = -(aua^\top + (a^\top)^{-1}ba^\top)$ and $\tilde{v} = \tilde{w}^2$ and the result follows by simple calculations. □

1.5.2 Study of the MLE of b with the Laplace transform

We consider $\epsilon : \mathbb{R}_+ \to \mathbb{R}_+^*$ a (deterministic) decreasing function such that $\lim_{t\to+\infty}\epsilon_t = 0$. From the definition of the MLE of b Equation (1.14), we get that

$$\frac{1}{\epsilon_T}(\hat{b}_T - b) = \mathcal{L}^{-1}_{\epsilon_T^2 R_T}(\epsilon_T[X_T - x - \alpha T I_d - bR_T - R_T b]).$$

Thus, we want to calculate the Laplace transform of $(\epsilon_T[X_T - x - \alpha T I_d - bR_T - R_T b], \epsilon_T^2 R_T)$ in order to study the convergence of $\frac{1}{\epsilon_T}(\hat{b}_T - b)$. For $\lambda_1, \lambda_2 \in \mathcal{S}_d$, we define

$$\mathcal{E}(T, \lambda_1, \lambda_2) := \mathbb{E}_\theta\Big[\exp\Big(-\epsilon_T\operatorname{Tr}[\lambda_2(X_T - x - \alpha T I_d - bR_T - R_T b)] - \epsilon_T^2\operatorname{Tr}[\lambda_1 R_T]\Big)\Big] \tag{1.53}$$

$$= \exp\left(\epsilon_T\operatorname{Tr}[\lambda_2(x + \alpha T I_d)]\right)\mathbb{E}_\theta\Big[\exp\Big(-\operatorname{Tr}[\epsilon_T\lambda_2 X_T] - \operatorname{Tr}[(\epsilon_T^2\lambda_1 - \epsilon_T(\lambda_2 b + b\lambda_2))R_T]\Big)\Big]. \tag{1.54}$$

We now consider $\lambda_1, \lambda_2 \in \mathcal{S}_d$ such that

$$\lambda_1 - 2\lambda_2^2 \in \mathcal{S}_d^{+,*}. \tag{1.55}$$

We define

$$v_T = 2\lambda_1\epsilon_T^2 - 2(b\lambda_2 + \lambda_2 b)\epsilon_T, \quad \tilde{v}_T = v_T + b^2, \quad w_T = 2\lambda_2\epsilon_T, \quad \tilde{w}_T = w_T - b, \tag{1.56}$$

and have $v_T + bw_T + w_T b - w_T^2 = \epsilon_T^2(2\lambda_1 - 4\lambda_2^2) \in \mathcal{S}_d^{+,*}$. Thus, by applying Proposition 1.5.1 with $m = -\epsilon_T\lambda_2$, we get that $\mathcal{E}(T, \lambda_1, \lambda_2)$ is finite and given by

$$\begin{aligned}\mathcal{E}(T, \lambda_1, \lambda_2) =& \frac{\exp\left(-\frac{\alpha}{2}\operatorname{Tr}[b]T\right)}{\det[V_{v_T,w_T}(T)]^{\frac{\alpha}{2}}}\exp\Big(-\frac{1}{2}\operatorname{Tr}\left[(V'_{v_T,w_T}(T)V_{v_T,w_T}(T)^{-1} + b)x\right]\Big)\\&\times\exp\Big(\epsilon_T\operatorname{Tr}\left[\lambda_2(x + \alpha T I_d)\right]\Big)\end{aligned} \tag{1.57}$$

with

$$V_{v_T,w_T}(T) =(\sqrt{\tilde{v}_T})^{-1}\sinh(\sqrt{\tilde{v}_T}T)\tilde{w}_T + \cosh(\sqrt{\tilde{v}_T}T)$$
$$V'_{v_T,w_T}(T) =\cosh(\sqrt{\tilde{v}_T}T)\tilde{w}_T + \sinh(\sqrt{\tilde{v}_T}T)\sqrt{\tilde{v}_T}.$$

Besides, we have $\tilde{v}_T = (b-2\epsilon_T\lambda_2)^2 + \epsilon_T^2(2\lambda_1 - 4\lambda_2^2) \in \mathcal{S}_d^{+,*}$.
When $-b \in \mathcal{S}_d^{+,*}$ and $\epsilon_T = 1/\sqrt{T}$, we can make explicit calculations and get

$$\lim_{T\to+\infty}\mathcal{E}(T,\lambda_1,\lambda_2) = \exp(-\operatorname{Tr}[\lambda_1\overline{R}_\infty] - \operatorname{Tr}[2\lambda_2^2\overline{R}_\infty]),$$

which gives another mean to prove Theorem 1.2.3. Here, we prove Theorem 1.3.4.

Proof of Theorem 1.3.4. Here, we focus on the case $b = b_0 I_d$ with $b_0 > 0$ and set $\epsilon_T = e^{-b_0 T}$. Since the square root function is analytic on the set of positive definite matrices (see e.g. [64], p. 134) we get that

$$\sqrt{\tilde{v}_T} = b_0 I_d - 2\epsilon_T\lambda_2 + \frac{\epsilon_T^2}{b_0}(\lambda_1 - 2\lambda_2^2) + O(\epsilon_T^3),$$

since the squares of each sides coincides up to a $O(\epsilon_T^3)$ term. We observe that $\tilde{w}_T = 2\epsilon_T\lambda_2 - b_0 I_d$, and thus $\sqrt{\tilde{v}_T} + \tilde{w}_T = \frac{\epsilon_T^2}{b_0}(\lambda_1 - 2\lambda_2^2) + O(\epsilon_T^3)$.
We now write

$$V_{v_T,w_T}(T) =(\sqrt{\tilde{v}_T})^{-1}\left[\frac{1}{2}\exp(\sqrt{\tilde{v}_T}T)(\sqrt{\tilde{v}_T}+\tilde{w}_T) + \frac{1}{2}\exp(-\sqrt{\tilde{v}_T}T)(\sqrt{\tilde{v}_T}-\tilde{w}_T)\right]$$
$$V'_{v_T,w_T}(T) =\frac{1}{2}\exp(\sqrt{\tilde{v}_T}T)(\sqrt{\tilde{v}_T}+\tilde{w}_T) + \frac{1}{2}\exp(-\sqrt{\tilde{v}_T}T)(\tilde{w}_T-\sqrt{\tilde{v}_T}).$$

Since $\epsilon_T\exp(\sqrt{\tilde{v}_T}T) \underset{T\to+\infty}{\to} I_d$, we get $\frac{1}{\epsilon_T}V_{v_T,w_T}(T) \underset{T\to+\infty}{\to} \frac{1}{b_0}\left[\frac{1}{2b_0}(\lambda_1 - 2\lambda_2^2) + b_0 I_d\right]$ and $\frac{1}{\epsilon_T}V'_{v_T,w_T}(T) \underset{T\to+\infty}{\to} \frac{1}{2b_0}(\lambda_1 - 2\lambda_2^2) - b_0 I_d$. This yields to

$$V'_{v_T,w_T}(T)V_{v_T,w_T}(T)^{-1} + b_0 I_d \underset{T\to+\infty}{\longrightarrow} (\lambda_1 - 2\lambda_2^2)\left(\frac{1}{2b_0}(\lambda_1 - 2\lambda_2^2) + b_0 I_d\right)^{-1}.$$

We also have $\frac{\exp\left(-\frac{\alpha}{2}\operatorname{Tr}[b_0 I_d]T\right)}{\det[V_{v_T,w_T}(T)]^{\frac{\alpha}{2}}} = \frac{1}{\det[\epsilon_T^{-1}V_{v_T,w_T}(T)]^{\frac{\alpha}{2}}} \underset{T\to+\infty}{\to} \frac{1}{\det\left[\frac{1}{b_0}\left[\frac{1}{2b_0}(\lambda_1-2\lambda_2^2)+b_0 I_d\right]\right]}$, and therefore

$$\lim_{T\to+\infty}\mathcal{E}(T,\lambda_1,\lambda_2) = \frac{\exp\left(-\frac{1}{2b_0}\operatorname{Tr}\left[(\lambda_1 - 2\lambda_2^2)\left(\frac{1}{2b_0^2}(\lambda_1 - 2\lambda_2^2) + I_d\right)^{-1}x\right]\right)}{\det\left[\frac{1}{2b_0^2}(\lambda_1 - 2\lambda_2^2) + I_d\right]}. \tag{1.58}$$

We now want to identify the limit. We know that $X \sim WIS_d\left(\frac{x}{2b_0}, \alpha, 0, I_d; \frac{1}{4b_0^2}\right)$ has the following Laplace transform

$$u \in \mathcal{S}_d^+, \mathbb{E}[\exp(-\operatorname{Tr}[uX])] = \frac{\exp\left(-\operatorname{Tr}\left[u(I_d + \frac{1}{2b_0^2}u)^{-1}\frac{x}{2b_0}\right]\right)}{\det[I_d + \frac{1}{2b_0^2}u]}.$$

Let $\tilde{\mathbf{G}}$ denote a d-square matrix independent from X, whose entries are independent and follow a standard Normal distribution. By Lemma 1.7.2, we have

$$\mathbb{E}[\exp(-\operatorname{Tr}[\lambda_1 X + \lambda_2(\sqrt{X}\tilde{\mathbf{G}} + \tilde{\mathbf{G}}\sqrt{X})])] = \mathbb{E}[\exp(-\operatorname{Tr}[(\lambda_1 - 2\lambda_2^2)X])].$$

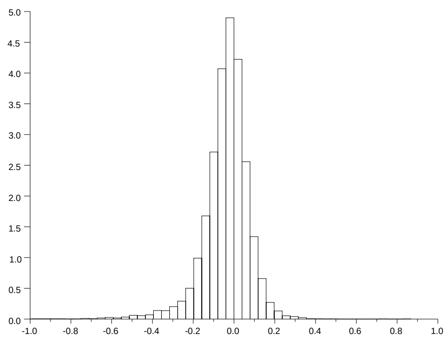
Thus, Equation (1.58) shows the convergence in law of $(\epsilon_T(X_T - x - \alpha T I_d - bR_T - R_T b), \epsilon_T^2 R_T)$ to $\left(X, \sqrt{X}\tilde{\mathbf{G}} + \tilde{\mathbf{G}}\sqrt{X}\right)$ under $\mathbb{P}_\theta$, which gives the claim of Theorem 1.3.4. □

1.6 Numerical illustration

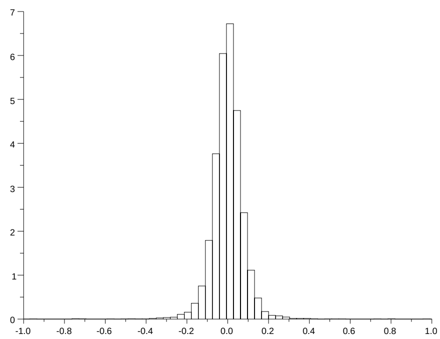
In this section, we test the convergence of the MLE given by Equation (1.12) and Equation (1.14). To do so, we consider a given large value of T and simulate the Wishart process exactly on the regular time grid $t_i = \frac{iT}{N}$, $i = 0, \cdots, N$. This can be done by using the method presented in Ahdida and Alfonsi [1]. We take N sufficiently large and approximate the integrals R_T and Q_T^{-1} applying the trapezoidal rule along this time grid. Thus, we will use the estimator with the exact value of X_T and these approximated values of R_T and Q_T^{-1}.
This section has three goals. First, we check numerically the convergence results that we have obtained. Second, we investigate numerically the convergence of the MLE in some nonergodic cases, where no theoretical result of convergence have been found. Last, we test the estimation of the parameters of a full Wishart process Equation (1.1). To do so, we estimate first a with the quadratic variation and then the parameters α and b by using the MLE Equation (1.12) on the process $(a^\top)^{-1}Xa^{-1}$.

1.6.1 Numerical validation of the convergence results

Using the method mentioned above, we have checked the convergence results obtained in this paper. Namely, we sample $M = 10000$ independent paths of X in order to draw an histogram of the properly rescaled value of $\hat{b}_{i,j} - b_{i,j}$ or $\hat{\alpha} - \alpha$. We do not reproduce all these graphics here, a present for example in Figure 1.1 an illustration of the convergence given by Theorem 1.3.4.



(a) Limit law of $\exp(0.05T)(b - \hat{b}_T)_{1,1}$.

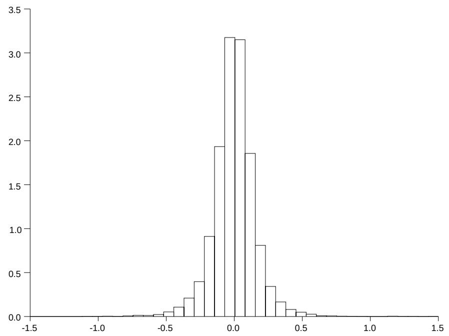


(b) Limit law of $\exp(0.05T)(b - \hat{b}_T)_{1,2}$.

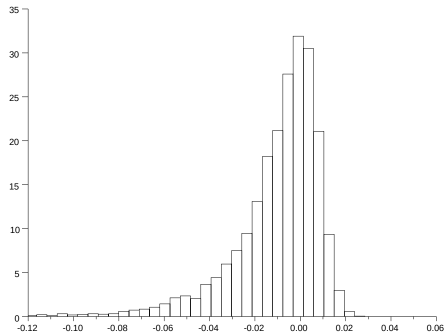
Figure 1.1: Asymptotic law of the error for the estimation of $\theta = b$ with for: $x = \left(\begin{smallmatrix} 0.5 & 0.1 \\ 0.1 & 0.3 \end{smallmatrix}\right)$, $T = 100$, $N = 10000$, $\alpha = 4.5$ and $b = 0.05 I_d$.

1.6.2 Experimental convergence in a nonergodic case

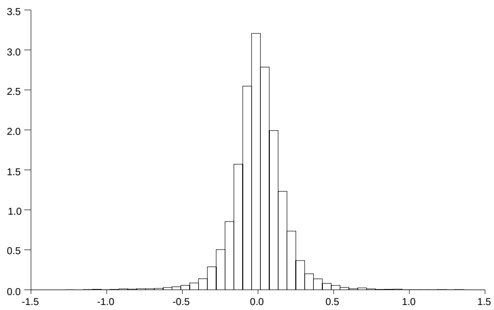
In this paragraph, we try to guess the asymptotic behavior of the MLE in an ergodic case, where no theoretical convergence result is known. Namely, we observe in Figure 1.2 the asymptotic estimation error, when $b = \mathrm{diag}(0.1, 0.005)$ is diagonal with positive and distinct terms on its diagonal and when we use the estimator Equation (1.14). As one might have guess, the convergence of the diagonal terms seems to be with an exponential rate, with the exponential speed corresponding to its value. Namely, $\hat{b}_{11}$ seems to converge to b_{11} with a speed of $\exp(0.1T)$ while $\hat{b}_{22}$ seems to converge to b_{22} with a speed of $\exp(0.005T)$. More interesting is the antidiagonal term. One could have imagine that the convergence rate is the slowest of these two rates. Instead, on our experiment, the convergence of $\hat{b}_{12}$ towards b_{12} seems to happen with the rate $\exp(0.1T)$. We have observed the same behaviour for other parameter values. Of course, it would be hasty to draw a global conclusion from few particular experiments. However, it is interesting to note that these numerical tests are a way to guess or check the convergence rate of the MLE.



(a) Limit law of $\exp(0.1T)(b - \hat{b}_T)_{1,1}$.



(b) Limit law of $\exp(0.005T)(b - \hat{b}_T)_{2,2}$.



(c) Limit law of $\exp(0.1T)(b - \hat{b}_T)_{1,2}$.

Figure 1.2: Asymptotic law of the error for the estimation of $\theta = b$ with $x = \left(\begin{smallmatrix} 0.3 & 0.1 \\ 0.1 & 0.2 \end{smallmatrix}\right)$, $T = 100$, $N = 10000$, $\alpha = 3.5$ and $b = \mathrm{diag}(0.1, 0.005)$.

1.6.3 Estimation of the whole Wishart process

In this last part of the numerical illustration, we perform the estimation of all the parameters of the Wishart process Equation (1.1). We consider a case where a is upper triangular and $(a^\top)^{-1}ba^\top$ is symmetric. We proceed as follows. First, we sample exactly a discrete path $(X_{iT/N}, 0 \leqslant i \leqslant N)$. Then, we estimate the matrix $a^\top a$ by using Equation (1.3), where the quadratic variations are replaced by their classical approximations and the integrals are replaced by the trapezoidal rule. By a Cholesky decomposition we get then an estimator $\hat{a}$ of a. Then, we use the MLE Equation (1.12) on the path $((\hat{a}^\top)^{-1}X_{iT/N}\hat{a}^\top, 0 \leqslant i \leqslant N)$. This gives an estimator of α and $(a^\top)^{-1}ba^\top$, and therefore an estimator of b. As a comparison, we also calculate similarly the estimator of α and b when a is known and has not to be estimated. To draw histograms or calculate empirical expectations, we run $M = 10000$ independent paths of X.
We consider a sufficiently large value of T and are interested in looking at the convergence with respect to N. First, we plot the the error on the estimator of a with respect to the number of time step in Log-Log scale. We observe that the convergence to zero takes place with experimental rate close to $1/2$. This is in line with the general results on the estimation of the diffusion

Number of time steps		20	50	100	200	500	1000	2000	5000
$\mathbb{E}[\mathrm{Tr}[(a-\widehat{a}^N)^2]]^{1/2}$		1.7671	1.4311	1.1487	0.7913	0.4107	0.2472	0.1514	0.0846
$\mathrm{MSE}(\widehat{b}^N_{1,1}\|b_{1,1})$	$\widehat{a}=a$	0.0745	0.0338	0.0181	0.0115	0.0082	0.0069	0.0061	0.0058
	$\widehat{a}=\widehat{a}^N$	0.7636	0.5266	0.3489	0.1891	0.0624	0.0273	0.0142	0.0085
$\mathrm{MSE}(\widehat{b}^N_{2,2}\|b_{2,2})$	$\widehat{a}=a$	0.2554	0.1310	0.0664	0.0372	0.0231	0.0176	0.0153	0.0139
	$\widehat{a}=\widehat{a}^N$	3.4085	2.8722	2.1159	1.1995	0.3600	0.1264	0.0480	0.0201
$\mathrm{MSE}(\widehat{b}^N_{1,2}\|b_{1,2})$	$\widehat{a}=a$	0.0075	0.0033	0.0017	0.0011	0.0008	0.0008	0.0007	0.0007
	$\widehat{a}=\widehat{a}^N$	0.0442	0.0568	0.0596	0.0352	0.0148	0.0075	0.0039	0.0019
$\mathrm{MSE}(\widehat{\alpha}^N\|\alpha)$	$\widehat{a}=a$	0.8448	0.3579	0.1993	0.1151	0.0614	0.0416	0.0308	0.0230
	$\widehat{a}=\widehat{a}^N$	0.8267	0.3496	0.1895	0.1095	0.0617	0.0410	0.0311	0.0234

Table 1.1: Mean Squared Error for the estimation of $\theta = (\alpha, b)$ with respect to N. Same parameters as Figure 1.3.

coefficient, see Dohnal [25] and Genon-Catalot and Jacod [27]. Then, we focus on the influence of the discretization and the unknown parameter a on the convergence of the MLE of b and α. In Table 1.1, we give in function of N the Mean Squared Error $\mathrm{MSE}(\widehat{\theta}^N|\theta) = \mathbb{E}[|\widehat{\theta}^N - \theta|^2]$ of the estimator $\widehat{\theta}^N$, with $\theta = (b, \alpha)$. It is estimated with the empirical expectation. First, we observe that the convergence of the estimator of α is roughly the same whether we know a or not. This is expected since the estimation of α does not depend on the estimation of a. Instead, the bias on b is much higher when a is estimated than when a is known. However, it decreases also faster at an experimental order of 0.7 while the bias when a is known decreases at an experimental order of 0.45. This latter rate is in line with the rate of 1/2 obtained in dimension 1 by Ben Alaya and Kebaier [14]. In our case, it seems that the influence of the estimation of a vanishes around $N = 5000$. Last, we have plotted in Figure 1.4 the limit law of the estimator $\sqrt{T}(\widehat{\theta}^N - \theta)$ with $N = 10000$.

This short numerical study shows that the estimator obtained by discretizing the continuous time estimator is efficient in practice. Of course, it would be nice to obtain general convergence results in function of T and N, but we leave this for further research.

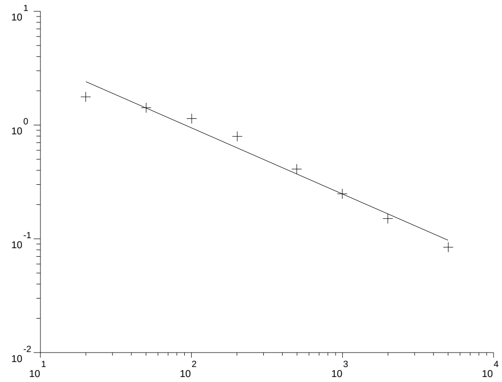
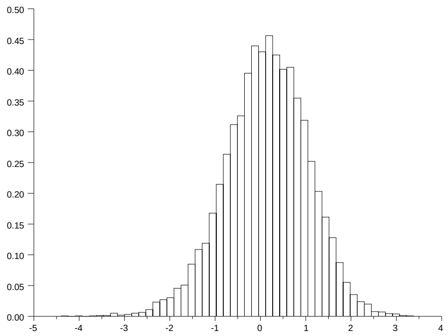
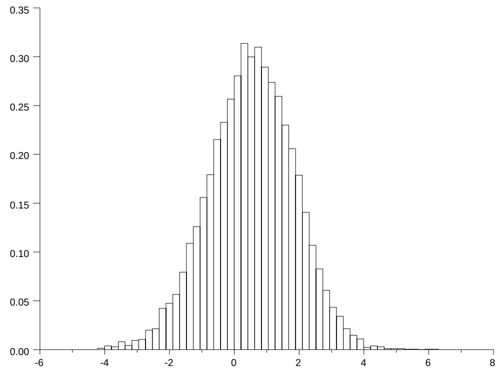


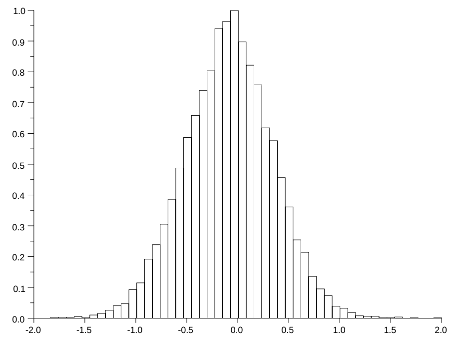
Figure 1.3: Log-Log representation of the empirical expectation of $\mathbb{E}[\mathrm{Tr}[(a - \widehat{a}^N)^2]]^{1/2}$ for $x = \left(\begin{smallmatrix} 0.8 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$, $T = 100$, $a = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 2 \end{smallmatrix}\right)$, $\alpha = 4.5$, $b = \begin{pmatrix} -1 & 0.2 \\ 2 & -2 \end{pmatrix}$, where the line is the simple linear regression i.e. $\log(\mathbb{E}[\mathrm{Tr}[(a - \widehat{a}^N)^2]]^{1/2}) \approx 2.62 - 0.58 \log(N)$.



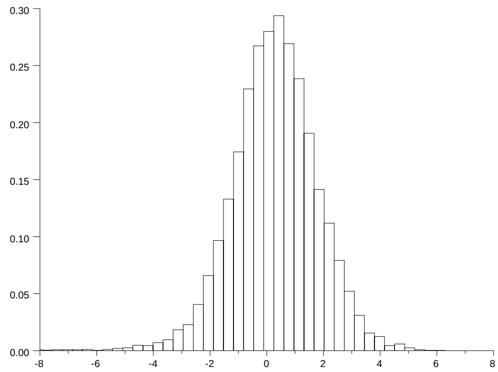
(a) Limit law of $\sqrt{T}(b - \hat{b}_T^N)_{1,1}$.



(b) Limit law of $\sqrt{T}(b - \hat{b}_T^N)_{2,2}$.



(c) Limit law of $\sqrt{T}(b - \hat{b}_T^N)_{1,2}$.



(d) Limit law of $\sqrt{T}(\alpha - \hat{\alpha}_T^N)$.

Figure 1.4: Asymptotic laws of the error for the estimation of $\theta = (\alpha, b)$ for $\widehat{a} = \widehat{a}^N$, $N = 10000$, same parameters as Figure 1.3.

1.7 Appendix

1.7.1 Proof of Proposition 1.1.1

We work under the probability $\mathbb{P}_{\theta_0,T}$ and have (see eq. (2.9) in Bru [20])

$$d\log(\det[X_t]) = (\alpha_0 - 1 - d)\operatorname{Tr}[X_t^{-1}]dt + 2\operatorname{Tr}[(\sqrt{X_t})^{-1}d\tilde{W}_t].$$

We denote $b^s = (b+b^\top)/2$ (resp. $b^a = (b-b^\top)/2$) the symmetric (resp. antisymmetric) part of b. We have

$$\int_0^T \operatorname{Tr}[(\sqrt{X_s})^{-1}d\tilde{W}_s] = \frac{1}{2}\log\left(\frac{\det[X_T]}{\det[x]}\right) - \frac{1}{2}\int_0^T (\alpha_0 - 1 - d)\operatorname{Tr}[X_s^{-1}]ds,$$

$$\begin{aligned}\int_0^T \operatorname{Tr}[b\sqrt{X_s}d\tilde{W}_s] &= \frac{1}{2}\int_0^T \operatorname{Tr}[b^s(\sqrt{X_s}d\tilde{W}_s + d\tilde{W}_s^\top\sqrt{X_s})] + \int_0^T \operatorname{Tr}[b^a\sqrt{X_s}d\tilde{W}_s]\\ &= \frac{\operatorname{Tr}[b^sX_T] - \operatorname{Tr}[b^sx]}{2} - \frac{\alpha_0 T}{2}\operatorname{Tr}[b^s] + \int_0^T \operatorname{Tr}[b^a\sqrt{X_s}d\tilde{W}_s].\end{aligned}$$

Therefore, we get from Equation (1.6) that $\frac{d\mathbb{P}_{\theta,T}}{d\mathbb{P}_{\theta_0,T}} \in \mathcal{F}_T^X$ if, and only if, $\int_0^T \operatorname{Tr}[b^a\sqrt{X_s}d\tilde{W}_s] \in \mathcal{F}_T^X$. We note that this latter variable is square integrable since X has bounded moments, and we set

$$M_t^a = \int_0^t \operatorname{Tr}[b^a\sqrt{X_s}d\tilde{W}_s].$$

If $\int_0^T \operatorname{Tr}[b^a\sqrt{X_s}d\tilde{W}_s] \in \mathcal{F}_T^X$, then we know by using martingale representation results (see e.g. Rogers and Williams [64], Theorem V.25.1) that there is a $\mathcal{F}^X$-previsible process $(U_t, t \in [0,T])$ taking values in $\mathcal{S}_d$, satisfying $\mathbb{E}[\int_0^T \operatorname{Tr}[U_tX_tU_t]dt] < \infty$ and such that

$$M_t^a = \int_0^t \operatorname{Tr}[U_s d(X_s - \alpha_0 I_d s)] = 2\int_0^t \operatorname{Tr}[U_s\sqrt{X_s}d\tilde{W}_s].$$

Since $\langle \operatorname{Tr}[A_s d\tilde{W}_s], \operatorname{Tr}[B_s d\tilde{W}_s]\rangle = \operatorname{Tr}[A_sB_s^\top]ds$ for adapted matrix valued processes A and B, we get

$$\langle M^a\rangle_t = -\int_0^t \operatorname{Tr}[b^aX_sb^a]ds = -2\int_0^t \operatorname{Tr}[U_sX_sb^a]ds = 4\int_0^t \operatorname{Tr}[U_sX_sU_s]ds.$$

This leads to $\operatorname{Tr}[(2U_t - b^a)X_t(2U_t + b^a)] = 0$ a.s., dt-a.e on $(0,T)$. Since the function $Y \mapsto \operatorname{Tr}[Y^\top Y]$ is a norm on $\mathcal{M}_d$, we get $\sqrt{X}_t(2U_t + b^a) = 0$ *a.s.* Since $X_t \in \mathcal{S}_d^{+,*}$ a.s. dt-a.e., we get $b^a = -2U_t$ and then $b^a = 0$ since $\mathcal{S}_d \cap \mathcal{A}_d = \{0\}$.

Remark 1.7.1. *Proposition 1.1.1 indicates that the MLE is well defined only when $b \in \mathcal{S}_d$, and one may be interested to estimate b without this assumption. A natural idea is to look at other estimators such as the method of moments, at least when the process is ergodic. Let us consider the case $b = -\lambda I_d + b^a$ with $\lambda > 0$ and $b^a \in \mathcal{A}_d$, and X the solution of Equation (1.4). Then, we calculate for $v \in \mathcal{S}_d^+$ (see, e.g. Proposition 4 in [1]),*

$$\begin{aligned}\mathbb{E}[\exp(-\operatorname{Tr}[vX_t])] &= \frac{\exp\left(-\operatorname{Tr}\left[v\left(I_d + 2\frac{1-e^{-2\lambda t}}{2\lambda}v\right)^{-1} e^{-2\lambda t}e^{tb^a}xe^{-tb^a}\right]\right)}{\det\left[I_d + 2\frac{1-e^{-2\lambda t}}{2\lambda}v\right]^{\alpha/2}}\\ &\underset{t\to+\infty}{\to} \frac{1}{\det\left[I_d + \frac{1}{\lambda}v\right]^{\alpha/2}}.\end{aligned}$$

Thus, X_t converges in law to $X_\infty \sim WIS_d(0,\alpha,0,I_d;1/(2\lambda))$ when $t\to+\infty$ and the stationary law does not depend on b^a. Instead, if we consider now the couple (X_t,X_{t+1}), we have for $v_1,v_2\in\mathcal{S}_d^+$,

$$\begin{aligned}
&\mathbb{E}[\exp(-\operatorname{Tr}[v_1X_t+v_2X_{t+1}])]\\
&=\frac{\mathbb{E}\left[\exp\left(-\operatorname{Tr}\left[v_1X_t+v_2\left(I_d+\frac{1-e^{-2\lambda}}{\lambda}v_2\right)^{-1}e^{-2\lambda}e^{b^a}X_te^{-b^a}\right]\right)\right]}{\det\left[I_d+\frac{1-e^{-2\lambda}}{\lambda}v_2\right]^{\alpha/2}}\\
&\underset{t\to+\infty}{\to}\frac{1}{\det\left[I_d+\frac{1-e^{-2\lambda}}{\lambda}v_2\right]^{\alpha/2}\times\det\left[I_d+\frac{1}{\lambda}\left(v_1+e^{-b^a}v_2\left(I_d+\frac{1-e^{-2\lambda}}{\lambda}v_2\right)^{-1}e^{-2\lambda}e^{b^a}\right)\right]^{\alpha/2}}.
\end{aligned}$$

This stationary law of (X_t,X_{t+1}) depends on b^a, and it is then possible in principle to estimate b^a by considering the limit of $\frac{1}{T}\int_0^T f(X_s,X_{s+1})ds$ for different functions $f:(\mathcal{S}_d^+)^2\to\mathbb{R}$. We leave the estimation of $b\in\mathcal{M}_d$ by the method of moments for further research and focus on the MLE in this paper.

1.7.2 Technical lemmas

Lemma 1.7.1. *For $X\in\mathcal{S}_d^{+,*}$ and $a\geqslant 0$, let $\mathcal{L}_{X,a}$ and $\mathcal{L}_X=\mathcal{L}_{X,0}$ be the linear applications defined by Equation (1.9) on $\mathcal{S}_d$. If $a\operatorname{Tr}[X^{-1}]\neq 1$, then $\mathcal{L}_{X,a}$ is invertible and we have $\operatorname{Tr}[\mathcal{L}_{X,a}^{-1}(Y)]=\frac{\operatorname{Tr}[X^{-1}Y]}{2(1-a\operatorname{Tr}[X^{-1}])}$. Besides, the map $(X,Y,a)\mapsto\mathcal{L}_{X,a}^{-1}(Y)$ is continuous on $\{(X,Y,a)\in\mathcal{S}_d^{+,*}\times\mathcal{S}_d\times\mathbb{R}_+,\ a\operatorname{Tr}[X^{-1}]\neq 1\}$.*

Proof. The invertibility of $\mathcal{L}_{X,a}$ is equivalent to its one-to-one property, and we have

$$Y\in\ker(\mathcal{L}_{X,a})\iff YX+XY=2a\operatorname{Tr}[Y]I_d\Leftrightarrow X=2a\operatorname{Tr}[Y]Y^{-1}-YXY^{-1}$$

Since $X\in\mathcal{S}_d^{+,*}$, there exists an orthogonal matrix O_X and a diagonal matrix D_X with positive elements such that $X=O_XD_XO_X^\top$. We get

$$\begin{aligned}
Y\in\ker(\mathcal{L}_{X,a})&\iff O_XD_XO_X^\top=2a\operatorname{Tr}[Y]Y^{-1}-YO_XD_XO_X^\top Y^{-1}\\
&\iff D_X=2a\operatorname{Tr}[Y]O_X^\top Y^{-1}O_X-(O_X^\top YO_X)D_X(O_X^\top YO_X)^{-1}\\
&\iff D_X(O_X^\top YO_X)=2a\operatorname{Tr}[Y]I_d-(O_X^\top YO_X)D_X. \qquad (1.59)
\end{aligned}$$

Since D_X is diagonal, we obtain for $1\leqslant i,k\leqslant d$, $\Big((O_X^\top YO_X)D_X\Big)_{i,k}=(O_X^\top YO_X)_{i,k}(D_X)_{k,k}$ and $\Big(D_X(O_X^\top YO_X)\Big)_{i,k}=(D_X)_{i,i}(O_X^\top YO_X)_{i,k}$. For $k\neq i$, Equation (1.59) gives $(O_X^\top YO_X)_{i,k}=0$. For $k=i$, we get $(O_X^\top YO_X)_{i,i}(D_X)_{i,i}=a\operatorname{Tr}[Y]$ and therefore

$$\operatorname{Tr}[Y]=\operatorname{Tr}[O_X^\top YO_X]=\operatorname{Tr}[Y]a\sum_{i=1}^d\frac{1}{(D_X)_{i,i}}=\operatorname{Tr}[Y]a\operatorname{Tr}[X^{-1}].$$

Since $a\operatorname{Tr}[X^{-1}]\neq 1$, we obtain $\operatorname{Tr}[Y]=0$ and then $(O_X^\top YO_X)_{i,i}=0$, which gives $Y=0$ and the invertibility of $\mathcal{L}_{X,a}$. Let $c=\mathcal{L}_{X,a}^{-1}(Y)$. We have $c+XcX^{-1}-2a\operatorname{Tr}[c]X^{-1}=X^{-1}Y$, which gives $2(1-a\operatorname{Tr}[X^{-1}])\operatorname{Tr}[c]=\operatorname{Tr}[X^{-1}Y]$. Last, the continuity property is obvious since $(X,a)\mapsto\mathcal{L}_{X,a}$ is continuous and $\mathcal{L}\mapsto\mathcal{L}^{-1}$ is continuous on $\{\mathcal{L}:\mathcal{S}_d\to\mathcal{S}_d$ linear and invertible$\}$. □

The following lemma gives the Laplace transform of the matrix Normal distribution.

Lemma 1.7.2. *Let $C \in \mathcal{S}_d^{+,*}$ and $\mathcal{C}[C] \in (\mathbb{R}^d)^{\otimes 4}$ defined by*

$$\mathcal{C}[C]_{i,j,k,l} = \delta_{ik} C_{j,l} + \delta_{il} C_{j,k} + \delta_{jk} C_{i,l} + \delta_{jl} C_{i,k}. \tag{1.60}$$

We introduce the $\mathcal{M}_d$-valued random variables $\tilde{\mathbf{G}}$ and $\mathbf{G} \sim \mathcal{N}(0, \mathcal{C}[C])$ of which components are Normal random variables with mean 0 such that

$$\forall i,j,k,l \in \{1,\dots,d\}, \quad \mathbb{E}[\tilde{\mathbf{G}}_{i,j} \tilde{\mathbf{G}}_{k,l}] = \delta_{ik}\delta_{jl}, \quad \mathbb{E}[\mathbf{G}_{i,j}\mathbf{G}_{k,l}] = \mathcal{C}[C]_{i,j,k,l}. \tag{1.61}$$

We have the following results.

1. *For all $c \in \mathcal{S}_d$, $\mathbb{E}\big[\exp(-\operatorname{Tr}[c\mathbf{G}])\big] = \exp(2\operatorname{Tr}[c^2 C])$.*
2. *For $\tilde{C} \in \mathcal{M}_d$ such that $\tilde{C}\tilde{C}^\top = C$, $\tilde{C}\tilde{\mathbf{G}} + \tilde{\mathbf{G}}^\top \tilde{C}^\top$ and $\mathbf{G}$ have the same law.*
3. *Let $X \in \mathcal{S}_d^{+,*}$. For $c \in \mathcal{S}_d$, $\mathbb{E}[\exp(-\operatorname{Tr}[c\mathcal{L}_X^{-1}(\sqrt{X}\tilde{\mathbf{G}} + \tilde{\mathbf{G}}^\top\sqrt{X})])] = \mathbb{E}[\exp(\operatorname{Tr}[c\mathcal{L}_X^{-1}(c)])$*

Proof. We focus on the first point. For all $c \in \mathcal{S}_d$, we have

$$\mathbb{E}\big[\exp(-\operatorname{Tr}[c\mathbf{G}])\big] = \mathbb{E}\big[\exp(-\sum_{1\leqslant i,j\leqslant d} c_{i,j}\mathbf{G}_{i,j})\big].$$

Moreover, $\sum_{1\leqslant i,j\leqslant d} c_{i,j}\mathbf{G}_{i,j}$ is a Normal random variable and its variance is given by

$$\mathbb{E}\big[\big(\sum_{1\leqslant i,j\leqslant d} c_{i,j}\mathbf{G}_{i,j}\big)^2\big] = \sum_{1\leqslant i,j,k,l\leqslant d} c_{i,j}c_{k,l}\mathcal{C}[C]_{i,j,k,l} = 4\operatorname{Tr}[c^2C].$$

It follows from the moment generating function of the Normal distribution that

$$\mathbb{E}\big[\exp(-\operatorname{Tr}[c\mathbf{G}])\big] = \exp(2\operatorname{Tr}[c^2C]).$$

To prove the second point it is sufficient to notice that $\operatorname{Tr}[c(\tilde{C}\tilde{\mathbf{G}} + \tilde{\mathbf{G}}^\top\tilde{C}^\top)] = \operatorname{Tr}[2c\tilde{C}\tilde{\mathbf{G}}]$ and

$$\mathbb{E}\big[\big(\sum_{1\leqslant i,j\leqslant d} (c\tilde{C})_{i,j}\tilde{\mathbf{G}}_{i,j}\big)^2\big] = \sum_{1\leqslant i,j,k,l\leqslant d} (c\tilde{C})_{i,j}(c\tilde{C})_{k,l}\delta_{ik}\delta_{jl} = \sum_{1\leqslant i,j\leqslant d} (c\tilde{C})^2_{i,j} = \operatorname{Tr}[c\tilde{C}\tilde{C}^\top c].$$

For the third point, we set $Z = \mathcal{L}_X^{-1}(\sqrt{X}\tilde{\mathbf{G}} + \tilde{\mathbf{G}}^\top\sqrt{X})$ and have $XZ + ZX = \sqrt{X}\tilde{\mathbf{G}} + \tilde{\mathbf{G}}^\top\sqrt{X}$. We also introduce $\tilde{c} = \mathcal{L}_X^{-1}(c)$ and have $\tilde{c}X + X\tilde{c} = c$. Thus, we obtain

$$\operatorname{Tr}[cZ] = \operatorname{Tr}[(\tilde{c}X + X\tilde{c})Z] = \operatorname{Tr}[\tilde{c}(\sqrt{X}\tilde{\mathbf{G}} + \tilde{\mathbf{G}}^\top\sqrt{X})]$$

and therefore $\mathbb{E}[\exp(-\operatorname{Tr}[cZ])] = \exp(2\operatorname{Tr}[\tilde{c}^2X]) = \exp(\operatorname{Tr}[\tilde{c}(\tilde{c}X + X\tilde{c})]) = \exp(\operatorname{Tr}[\tilde{c}c])$.

□

1.7.3 Some asymptotic behaviour of Wishart processes

Lemma 1.7.3. *Let $X \sim WIS_d(x, \alpha, b, I_d)$ with $b \in \mathcal{S}_d$, $x \in \mathcal{S}_d^+$ and $\alpha \geqslant d-1$. Then X_T converges in law when $T \to +\infty$ if and only if $-b \in \mathcal{S}_d^{+,*}$. In this case, X_T converges in law to $WIS_d(0, \alpha, 0, \sqrt{b^{-1}}; 1/2)$.*

Proof. Let us first consider the case $-b \in \mathcal{S}_d^{+,*}$. From Proposition 4 in [2], we have for $v \in \mathcal{S}_d^+$,

$$\mathbb{E}[\exp(-\operatorname{Tr}[vX_T])] = \frac{\exp\left(\operatorname{Tr}\left[-v\left(I_d + 2\left(\int_0^T e^{2bs}ds\right)v\right)^{-1}\right]e^{Tb}xe^{Tb}\right)}{\det\left[I_d - 2\left(\int_0^T e^{2bs}ds\right)v\right]^{\alpha/2}}$$
$$\underset{T\to+\infty}{\to} \frac{1}{\det\left[I_d - b^{-1}v\right]^{\alpha/2}},$$

which is the Laplace transform of $WIS_d(0, \alpha, 0, \sqrt{b^{-1}}; 1/2)$. Now, let us consider $-b \notin \mathcal{S}_d^{+,*}$. Then, there exists an eigenvector $v \in \mathbb{R}^d \setminus \{0\}$ such that $bv = \lambda v$ with $\lambda \geqslant 0$. Then, we have $\frac{d}{dt}\mathbb{E}[v^\top X_t v] = \alpha v^\top v + 2\lambda \mathbb{E}[v^\top X_t v]$, and therefore $\mathbb{E}[v^\top X_T v] \underset{T\to+\infty}{\to} +\infty$. $\square$

Lemma 1.7.4. • *Assume $\alpha > d+1$ and $b = 0$. Then, $\frac{Q_T^{-1}}{d\log(T)} \underset{T\to+\infty}{\to} \frac{1}{\alpha-(d+1)}$ a.s. Besides, $\frac{Z_T}{\log(T)}$ converges almost surely to d, and we have*

$$\forall \mu > 0, \sup_{T\geqslant 2} \mathbb{E}\left[\exp\left(\frac{\mu}{\sqrt{\log(T)}}N_T\right)\right] < \infty. \tag{1.62}$$

• *Assume $\alpha = d+1$ and $b = 0$. Then, as $T \to +\infty$, $\left(\frac{2}{d\log(T)}\right)^2 Q_T^{-1}$ converges in law to $\tau_1 = \inf\{t \geqslant 0, B_t = 1\}$, where B is a Brownian motion. Besides, $\frac{Z_T}{\log(T)} = \frac{2N_T}{\log(T)}$ converges in probability to d, and we have*

$$\forall \mu > 0, \sup_{T\geqslant 2} \mathbb{E}\left[\exp\left(\frac{\mu}{\log(T)}N_T\right)\right] < \infty. \tag{1.63}$$

We mention that the results on the convergence for Q_T are given in Donati-Martin et al. [26]. However, their proofs is in a working paper by the same authors that we have not been able to find. For this reason, we present here an autonomous proof.

Proof. We first consider the case $\alpha > d+1$. We have $dX_t = \alpha I_d dt + \sqrt{X_t}dW_t + dW_t^\top\sqrt{X_t}$ and thus

$$d(e^{-t}X_{e^t-1}) = [\alpha I_d - e^{-t}X_{e^t-1}]dt + \sqrt{e^{-t}X_{e^t-1}}d\tilde{W}_t + d\tilde{W}_t^\top\sqrt{e^{-t}X_{e^t-1}},$$

with $d\tilde{W}_t = e^{-t/2}d(W_{e^t-1})$. We observe that $\tilde{W}$ is a matrix Brownian motion, which gives $Y \sim WIS_d(x, \alpha, -I_d/2, I_d)$, where $Y_t = e^{-t}X_{e^t-1}$ for $t \geqslant 0$. Using equation Equation (1.21) to the process Y, we get

$$\frac{1}{t}\log\left(\frac{\det[Y_t]}{\det[Y_0]}\right) = (\alpha - 1 - d)\frac{1}{t}\int_0^t \operatorname{Tr}[Y_s^{-1}]ds - d + \frac{2}{t}\int_0^t \operatorname{Tr}[\sqrt{Y_s^{-1}}d\tilde{W}_s]. \tag{1.64}$$

Since Y is ergodic and $\langle\int_0^t \operatorname{Tr}[\sqrt{Y_s^{-1}}d\tilde{W}_s]\rangle = \int_0^t \operatorname{Tr}[Y_s^{-1}]ds$, we get that the left hand side converges in probability to zero and the right hand side converges a.s. to $(\alpha-1-d)\mathbb{E}[\operatorname{Tr}[Y_\infty^{-1}]] - d$, where $Y_\infty \sim WIS_d(0,\alpha,0,\sqrt{2}I_d;1/2)$ is the stationary law of Y. Therefore, $\frac{1}{t}\log\left(\frac{\det[Y_t]}{\det[Y_0]}\right)$ converges a.s. to zero. Since $\frac{1}{t}\log\left(\frac{\det[Y_t]}{\det[Y_0]}\right) = \frac{1}{t}\log\left(\frac{\det[e^{-t}X_{e^t-1}]}{\det[x]}\right) = \frac{1}{t}\log\left(\frac{\det[X_{e^t-1}]}{\det[x]}\right) - d$, we get that $\frac{Z_T}{\log(T)} = \frac{1}{\log(T)}\log\left(\frac{\det[X_T]}{\det[x]}\right)$ converges a.s. to d when $T\to+\infty$.

Now, we use Equation (1.21) taken at time $T = e^t - 1$ and Dubins-Schwarz theorem: there is a Brownian motion β such that for all $t \geqslant 0$,

$$\frac{\alpha-(1+d)}{Q_{e^t-1}t} + \frac{2\beta_{Q^{-1}_{e^t-1}}}{t} = \frac{1}{t}\log\left(\frac{\det[X_{e^t-1}]}{\det[x]}\right).$$

This gives that $\frac{\alpha-(1+d)}{Q_{e^t-1}t} \underset{t\to+\infty}{\to} d$ a.s., and therefore $\frac{Q_T^{-1}}{d\log(T)} \underset{T\to+\infty}{\to} \frac{1}{\alpha-(d+1)}$, a.s.

It remains to prove Equation (1.62). From Equation (1.21), we have $N_T = \frac{Z_T}{2} - \frac{\alpha-1-d}{2}Q_T^{-1} \leq \frac{Z_T}{2}$ and thus $\mathbb{E}\left[\exp\left(\frac{\mu}{\sqrt{\log(T)}}N_T\right)\right] \leqslant \mathbb{E}\left[\left(\frac{\det[X_T]}{\det[x]}\right)^{\frac{\mu}{2\sqrt{\log(T)}}}\right] < \infty$, since the moments of X are bounded. Again we set $t = \log(T+1)$, and for $\Lambda \in [0,1]$, we have from Equation (1.64)

$$\begin{aligned}
N_T &= \int_0^T \mathrm{Tr}[\sqrt{X_s^{-1}}dW_s] = \int_0^t \mathrm{Tr}[\sqrt{Y_s^{-1}}d\tilde{W}_s]\\
&= \Lambda\int_0^t \mathrm{Tr}[\sqrt{Y_s^{-1}}d\tilde{W}_s] + (1-\Lambda)\left(\frac{1}{2}\log\left(\frac{\det[Y_t]}{\det[x]}\right) + \frac{d}{2}t - \frac{\alpha-1-d}{2}\int_0^t \mathrm{Tr}[Y_s^{-1}]ds\right).
\end{aligned}$$

By Cauchy-Schwarz inequality, we get

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\frac{\mu}{\sqrt{\log(T+1)}}N_T\right)\right]\\
\leq& e^{\frac{\mu d(1-\Lambda)}{2}\sqrt{t}}\mathbb{E}^{\frac{1}{2}}\left[\left(\frac{\det[Y_t]}{\det[x]}\right)^{(1-\Lambda)\frac{\mu}{\sqrt{t}}}\right]\\
&\times \mathbb{E}^{\frac{1}{2}}\left[\exp\left(\frac{2\mu\Lambda}{\sqrt{t}}\int_0^t \mathrm{Tr}[\sqrt{Y_s^{-1}}d\tilde{W}_s] - \mu(1-\Lambda)\frac{\alpha-1-d}{\sqrt{t}}\int_0^t \mathrm{Tr}[Y_s^{-1}]ds\right)\right].
\end{aligned}$$

We now take $\Lambda = \Lambda_t = \frac{1}{2\epsilon_t}\left(-1+\sqrt{1+4\epsilon_t}\right)$ with $\epsilon_t = \frac{2\mu}{(\alpha-1-d)\sqrt{t}}$ in order to obtain $\frac{1}{2}\left(\frac{2\mu\Lambda_t}{\sqrt{t}}\right)^2 = \mu(1-\Lambda_t)\frac{\alpha-1-d}{\sqrt{t}}$. We note that for t large enough, $\Lambda_t \in [0,1]$. Besides, we have $\Lambda_t \underset{t\to+\infty}{=} 1-\epsilon_t + o(1/t)$, so that $\sqrt{t}(1-\Lambda_t)$ converges to $\frac{2\mu}{\alpha-1-d}$. From Theorem 4.1 in [54], the second expectation is then equal to 1, while the first one is bounded since Y is ergodic. This yields to Equation (1.62).

We now consider the case $\alpha = d+1$. We set again $t = \log(1+T)$ and have $T = e^t - 1$. Thus,

$$Z_T = \log\left(\frac{\det[X_T]}{\det[x]}\right) = \log\left(\frac{\det[e^tY_t]}{\det[x]}\right) = \log\left(\frac{\det[Y_t]}{\det[x]}\right) + dt.$$

Again, Y_t converges in law towards $WIS_d(0,\alpha,0\sqrt{2}I_d;1/2)$. Therefore, the ergodic theorem gives that $\frac{1}{t}\log\left(\frac{\det[Y_t]}{\det[x]}\right)$ converges in probability to 0, which yields to the convergence in probability of $\frac{Z_T}{\log(T)}$ to d. We now turn to the convergence of $\left(\frac{2}{d\log(T)}\right)^2 Q_T^{-1}$. We know from Theorem 4.1 in Mayerhofer [54] that for $T > 0$ and $\lambda \geqslant 0$,

$$\mathbb{E}\left[\exp\left(\frac{2\lambda}{d\log(1+T)}N_T - \frac{(2\lambda)^2}{2d^2\log(1+T)^2}Q_T^{-1}\right)\right] = 1.$$

From Equation (1.21), we have $N_T = Z_T/2$ and we write

$$\begin{aligned} 1 =& \mathbb{E}\left[\exp\left(\lambda - \frac{(2\lambda)^2}{2d^2\log(1+T)^2}Q_T^{-1}\right)\right] \\ &+ \mathbb{E}\left[\exp\left(-\frac{(2\lambda)^2}{2d^2\log(1+T)^2}Q_T^{-1}\right)\left(\exp\left(\frac{2\lambda}{d\log(1+T)}N_T\right) - \exp(\lambda)\right)\right] \end{aligned}$$

We now observe that $\exp\left(-\frac{(2\lambda)^2}{2d^2\log(1+T)^2}Q_T^{-1}\right) \leqslant 1$ and that

$$\mathbb{E}\left[\exp\left(\frac{2\lambda}{d\log(1+T)}N_T\right)\right] = \mathbb{E}\left[\left(\frac{\det[X_T]}{\det[x]}\right)^{\frac{\lambda}{d\log(1+T)}}\right] = e^{\lambda}\mathbb{E}\left[\left(\frac{\det[Y_t]}{\det[x]}\right)^{\frac{\lambda}{dt}}\right].$$

Since Y has bounded moments and is stationary, $\sup_{t\geqslant 1}\mathbb{E}\left[\left(\frac{\det[Y_t]}{\det[x]}\right)^{\frac{\lambda}{dt}}\right] < \infty$. This gives the uniform integrability Equation (1.63) and that

$$\mathbb{E}\left[\exp\left(-\frac{(2\lambda)^2}{2d^2\log(1+T)^2}Q_T^{-1}\right)\left(\exp\left(\frac{2\lambda}{d\log(1+T)}N_T\right) - \exp(\lambda)\right)\right] \underset{T\to+\infty}{\to} 0.$$

Therefore, $\lim_{T\to+\infty}\mathbb{E}\left[\exp\left(\lambda - \frac{(2\lambda)^2}{2d^2\log(1+T)^2}Q_T^{-1}\right)\right] = 1$, which gives the desired convergence in law. $\square$

Chapter 2

Complementary numerical study

In this Chapter, we propose a complementary numerical study to the one given in Part III Chapter 1. We first describe the Algorithm we have used in the simulation and then we focus on some numerical convergences results for some non ergodic cases we have not treated in that Chapter.

2.1 Simulation of the path of the Wishart process

In order to check numerically the speed of convergence of the MLE estimator and to observe the limit laws, we recall that we have to simulate a large number of paths of the Wishart process (and not only a realization at a given time T). Indeed our calculus of the MLEs involve some integrals related to the Wishart process namely R_T and Q_T (see 1.10). However, there is no simulation method for those paths and we simulate the Wishart process along a time grid. We describe briefly the method we adopt in this study. First of all, we give in Algorithm 1, the algorithm we have used in order to simulate the Wishart process on the time grid $t_i = iT/N$. This algorithm has been introduced in [1] and provides a realisation of a random variable with law $WIS_d(x, \alpha, b, I_d; t)$. In order to simulate the process along the time grid $\{t_i, i = 0, ., N\}$, we have thus used this algorithm N times with $t = T/N$ updating the initial value at each step. We recall that in our specific case, $a = I_d$ and that we estimate the integrals Q_T^{-1} and R_T using the trapezoidal rule on this time grid.

2.2 Complementary numerical study for some non ergodic cases

Using the method mentioned above, we approximate the process (X_T, R_T, Q_T) along the time grid $t_i = \frac{iT}{N}$, $i = 0, \cdots, N$ and assume that $a = I_d$ is known. For each example, we run $M = 10000$ independant paths of the Wishart process on this time grid, and we draw the asymptotic distributions of $\hat{b}_{i,j} - b_{i,j}$ and $\hat{\alpha} - \alpha$. We will focus on the estimation of α and b for some non ergodic cases we do not treat in Part III Chapter 1.

Case $b = 0$.

Through Figure 2.1, we observe the asymptotic behavior of the error for the estimation of the couple $\widehat{\theta}^N$, with $\theta = (b, \alpha)$ for $b = 0$ and $\alpha \geqslant d + 1$. For the convergence of the component b the rate is the same as soon as $\alpha \geqslant d + 1$. However for the estimator of α, we have to distinguish two cases : $\alpha = d + 1$ and $\alpha > d + 1$, which lead respectively to the convergence with speed $\log(T)$ and $\sqrt{\log(T)}$. The results of this numerical study are in line with what we exepcted from the theoritical study and we also remark that the estimation of the component b is faster

Algorithm 1 Exact simulation of $X_t \sim WIS_d(X_0, \alpha, b, I_d; t)$

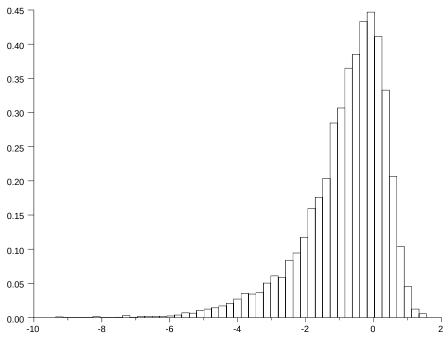
Input : $X_0 \in \mathcal{S}_d^+$, $\alpha > d-1$, $b \in \mathcal{S}_d$ and $t > 0$.
Output : $X_t \sim WIS_d(X_0, \alpha, b, I_d; t)$.
Set $q_t = \frac{1}{2} b^{-1}(\exp(2bt) - I_d)$.
Calculate (p_n, c_n, k_n) the extended Cholesky decomposition of $\frac{q_t}{t}$.
Set $\theta_t = p_n^{-1} \begin{pmatrix} c_n & 0 \\ k_n & I_{d-n} \end{pmatrix}$, $m_t = \exp(bt)$ and $y = \theta_t^{-1} m_t X_0 m_t (\theta_t^{-1})^\top$.
for $k = 1, \cdots, n$ **do**
 Set $p_{k,1} = p_{1,k} = p_{i,i}$ for $i \notin \{1, k\}$, $p_{i,j} = 0$ otherwise.
 Set $y = pyp$.
 Calculate (p_r, c_r, k_r) the extended Cholesky decomposition of $(y_{i,j})_{2 \leqslant i,j \leqslant d}$.
 Set $\pi = \begin{pmatrix} 1 & 0 \\ 0 & p_r \end{pmatrix}$, $\tilde{y} = \pi y \pi^\top$, $(u_{1,l+1})_{1 \leqslant l \leqslant r} = c_r^{-1}(\tilde{y}_{1,l+1})_{1 \leqslant l \leqslant r}$ and $u_{1,1} = \tilde{y}_{1,1} - \sum_{k=1}^r (u_{1,k+1})^2 \geqslant 0$.
 Sample independently $\mathbf{G_2}, \cdots, \mathbf{G_{r+1}} \sim \mathcal{N}(0,1)$ and the CIR starting from $u_{1,1}$ solving $d(U_t)_{1,1} = (\alpha - r)dt + 2\sqrt{(U_t)_{1,1}} dZ_t^1$.
 Set $(U_t)_{1,l+1} = u_{1,l+1} + \sqrt{t}\mathbf{G}_{l+1}$.
 Set

$$\begin{aligned} y =& \pi^\top \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_r & 0 \\ 0 & k_r & I_{d-r-1} \end{pmatrix} \\ &\times \begin{pmatrix} (U_t)_{1,1} - \sum_{k=1}^r ((U_t)_{1,k+1})^2 & ((U_t)_{1,l+1})_{1 \leqslant l \leqslant r}^\top & 0 \\ ((U_t)_{1,l+1})_{1 \leqslant l \leqslant r} & I_r & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &\times \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_r & k_r \\ 0 & 0 & I_{d-r-1} \end{pmatrix} \pi. \end{aligned}$$

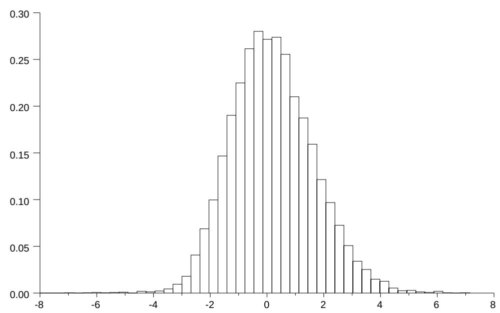
Return $X_t = \theta_t y \theta_t^\top$.

in this case whereas it is much more slower for α. Unfortunately, this slow convergence with rate $\log(T)$ or $\sqrt{\log(T)}$ can be an obstacle for concrete applications and it may be wise to use another estimator in this case.

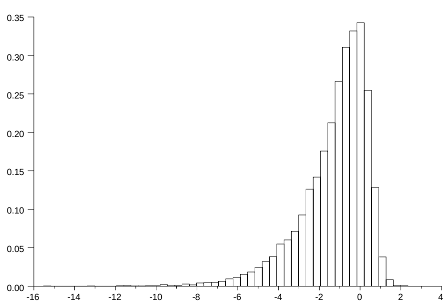
Remark 2.2.1. *In order to observe numerically the limit law of the error of the estimator α when $b = 0$, it requires a large number of realizations of the Wishart process for very long time. However, for very long horizon, such as $T = 10^{20}, 10^{50}$, the simulation of the Wishart process and of the integrals that appear in the estimator of α can be cumbersome. Indeed, in the one dimensional case, that is the CIR process, we can simulate exactly the law of (X_T, R_T, Q_T) which is no longer true for multi dimensional cases. We have to approximate the integrals on the time grid. However since T is large, it is quite impossible to consider a homogeneous time step that tends to zero for the approximation with reasonable time of calculus. The method we adopt here is ad hoc and consists in increasing the size of the time step as we get away from zero. This idea relies on the observation of Q_T. It is based on the fact that we are in a non ergodic case so that the approximation of Q_T seems particularly determined by the approximation of the part of the integral in the neighborhood of 0. In this case, this kind of time grid modification provides good numerical résults in reasonable time.*



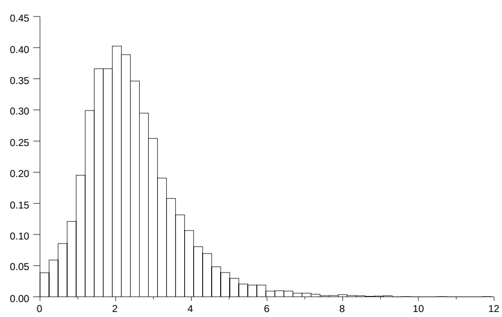
(a) Limit law of $T(b - \hat{b}_T^N)_{2,2}$.



(b) Limit law of $\sqrt{\log(T)}(\alpha - \hat{\alpha}_T^N)$.



(c) Limit law of $T(b - \hat{b}_T^N)_{1,1}$.



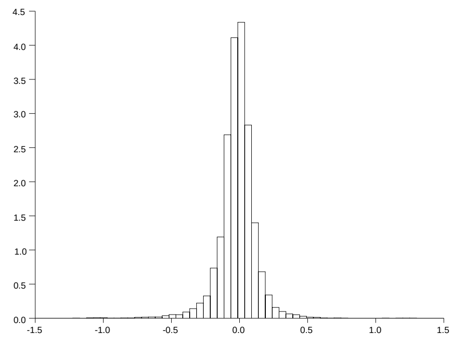
(d) Limit law of $\log(T)(\alpha - \hat{\alpha}_T^N)$.

Figure 2.1: Asymptotic laws of the error for the estimation of $\theta = (\alpha, b)$ for $X_0 = 0.3 I_d$, $b = 0$, and (from (a) to (d)) $(\alpha, T) = (4.5, 100), (4, 10^{50}), (3, 100), (3, 10^{20}), and N = 10000$.

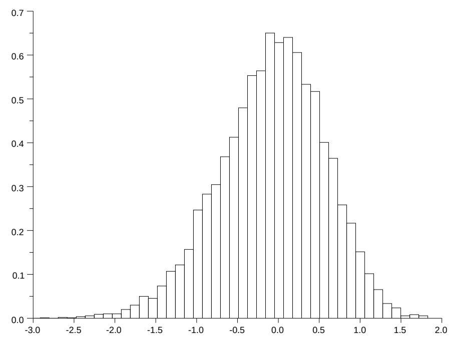
Case $b = \mathrm{diag}(b_1, ., b_d)$**.**

Finally, in Figure 2.2, we represent some examples of asymptotic error distributions of the estimator of b when b is known a priori diagonal and is given in (1.34). As we expected from the theory, for a given $i \in \{1,2\}$, $\widehat{b}_i^N$ converges with its own regime depending on the value of b_i. Each component behaves similarly as in the CIR case, which is largely explored in [14], and then it is possible to observe simultaneously both ergodic and non ergodic behaviors. The reader may wonder if the estimator is relevant to estimate the diagonal of b when b is not diagonal but it is not adapted in this case.

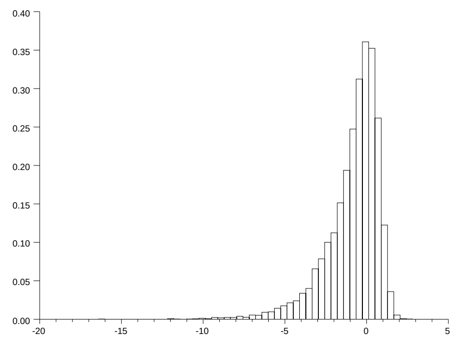
This study confirms that in a non ergodic case with disctinct eigenvalues for b, there appears miscellaneous speeds of convergence for the components of $\hat{b}_T - b$. Even if we are able to identify those speeds in the case where b is known a priori diagonal, the general case $b \in \mathcal{S}_d$ is much more difficult to treat because there is a mixing of all these types of speeds and we were not able to identify how they mix. The main difficulty relies thus on the fact that must consider not scalar but matricial normalization for $\hat{b}_T - b$ in order to identify the limit law.



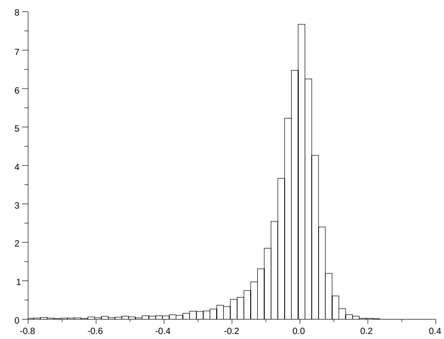
(a) Limit law of $\exp(0.06T)(b - \hat{b}_T)_{1,1}$.



(b) Limit law of $\sqrt{T}(b - \hat{b}_T)_{2,2}$.



(c) Limit law of $T(b - \hat{b}_T)_{1,1}$.



(d) Limit law of $\exp(0.03T)(b - \hat{b}_T)_{2,2}$.

Figure 2.2: Asymptotic law of the error for the estimation of $\theta = b$ with for (a), (b):$X_0 = \left(\begin{smallmatrix} 0.5 & 0.1 \\ 0.1 & 0.2 \end{smallmatrix}\right)$, $T = 100$, $\alpha = 2.5$ and $b = \mathrm{diag}(0.06, -0.5)$, and for (c), (d): $X_0 = \left(\begin{smallmatrix} 0.5 & 0.1 \\ 0.1 & 0.2 \end{smallmatrix}\right)$, $T = 100$, $\alpha = 1.5$ and $b = \mathrm{diag}(0, 0.03)$.

Bibliography

[1] A. Ahdida and A. Alfonsi. Exact and high-order discretization schemes for Wishart processes and their affine extensions. *Ann. Appl. Probab.*, 23(3):1025–1073, 2013. 25, 26, 27, 98, 116, 119, 123, 129

[2] A. Ahdida, A. Alfonsi, and E. Palidda. Smile with the Gaussian term structure model. *Working paper*, 2014. 26, 98, 126

[3] A. Alfonsi. High order discretization schemes for the CIR process: application to affine term structure and Heston models. *Math. Comp.*, 79(269):209–237, 2010. 14, 16, 22, 23, 37, 75, 83, 84, 85, 86, 87

[4] A. Alfonsi, A. Kebaier, and C. Rey. Maximum likelihood estimation for Wishart processes, Aug. 2015. 97

[5] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014. 18, 39

[6] V. Bally and L. Caramellino. On the distance between probability density functions. Nov. 2013. 18, 38, 50, 52

[7] V. Bally and L. Caramellino. Asymptotic development for the CLT in total variation distance. *ArXiv e-prints*, July 2014. 50

[8] V. Bally and L. Caramellino. Convergence and regularity of probability laws by using an interpolation method. *arXiv preprint arXiv:1409.3118*, 2014. 19, 39, 44, 46, 47

[9] V. Bally and E. Clément. Integration by parts formula and applications to equations with jumps. *Probab. Theory Related Fields*, 151(3-4):613–657, 2011. 15, 16, 18, 19, 38, 50, 52, 83, 85, 89

[10] V. Bally and C. Rey. Approximation of Markov semigroup in total variation disctance. Jan. 2015. 35, 83, 85, 87, 89, 94

[11] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function. *Probab. Theory Related Fields*, 104(1):43–60, 1996. 13, 15, 16, 37, 46, 85, 89

[12] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density. *Monte Carlo Methods Appl.*, 2(2):93–128, 1996. 13, 15, 16, 37, 85, 89

[13] M. Ben Alaya and A. Kebaier. Parameter estimation for the square-root diffusions: ergodic and nonergodic cases. *Stoch. Models*, 28(4):609–634, 2012. 25, 26, 29, 98, 100, 101, 109

[14] M. Ben Alaya and A. Kebaier. Asymptotic behavior of the maximum likelihood estimator for ergodic and nonergodic square-root diffusions. *Stoch. Anal. Appl.*, 31(4):552–573, 2013. 25, 26, 32, 98, 101, 109, 121, 132

[15] R. Bhattacharya and R. Rao. *Normal Approximation and Asymptotic Expansions.* Society for Industrial and Applied Mathematics, 2010. 36

[16] S. G. Bobkov, G. P. Chistyakov, and F. Götze. Berry–Esseen bounds in the entropic central limit theorem. *Probab. Theory Related Fields*, 159(3-4):435–478, 2014. 49

[17] S. G. Bobkov, G. P. Chistyakov, and F. Götze. Fisher information and the central limit theorem. *Probab. Theory Related Fields*, 159(1-2):1–59, 2014. 49

[18] M. Bossy, E. Gobet, and D. Talay. A symmetrized Euler scheme for an efficient approximation of reflected diffusions. *J. Appl. Probab.*, 41(3):877–889, 2004. 16, 37, 84

[19] M. Bru. *Thèse* $3^{ème}$ *cycle. Résistance d'Escherichia coli aux antibiotiques: Sensibilités des analyses en composantes principales aux perturbations Browniennes et simulation.* PhD thesis, Université Paris Nord, 1987. 25, 98

[20] M. Bru. Wishart processes. *J. Theoret. Probab.*, 4(4):725–751, 1991. 25, 27, 31, 97, 98, 100, 102, 112, 123

[21] C. Cuchiero, D. Filipović, E. Mayerhofer, and J. Teichmann. Affine processes on positive semidefinite matrices. *Ann. Appl. Probab.*, 21(2):397–463, 2011. 25, 97

[22] J. Da Fonseca, M. Grasselli, and F. Ielpo. Estimating the Wishart affine stochastic correlation model using the empirical characteristic function. *Stud. Nonlinear Dyn. Econom.*, 18(3):253–289, 2014. 26, 98

[23] J. Da Fonseca, M. Grasselli, and C. Tebaldi. Option pricing when correlations are stochastic: an analytical framework. *Review of Derivatives Research*, 10:151–180, 2008. 26, 98

[24] L. Dieci and T. Eirola. Positive definiteness in the numerical solution of Riccati differential equations. *Numer. Math.*, 67(3):303–313, 1994. 113

[25] G. Dohnal. On estimating the diffusion coefficient. *J. Appl. Probab.*, 24(1):105–114, 1987. 121

[26] C. Donati-Martin, Y. Doumerc, H. Matsumoto, and M. Yor. Some properties of the Wishart processes and a matrix extension of the Hartman-Watson laws. *Publ. Res. Inst. Math. Sci.*, 40(4):1385–1412, 2004. 126

[27] V. Genon-Catalot and J. Jacod. On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 29(1):119–151, 1993. 121

[28] A. Gnoatto. The Wishart short rate model. *Int. J. Theor. Appl. Finance*, 15(8):1250056, 24, 2012. 26, 98

[29] A. Gnoatto and M. Grasselli. The explicit Laplace transform for the Wishart process. *J. Appl. Probab.*, 51(3):640–656, 2014. 31, 97, 101, 112, 116, 117

[30] E. Gobet. Weak approximation of killed diffusion using Euler schemes. *Stochastic Process. Appl.*, 87(2):167–197, 2000. 16, 37, 84

[31] E. Gobet and S. Menozzi. Stopped diffusion processes: boundary corrections and overshoot. *Stochastic Process. Appl.*, 120(2):130–162, 2010. 16, 37, 84

[32] C. Gourieroux and R. Sufana. Derivative pricing with wishart multivariate stochastic volatility. *J. Bus. Econom. Statist.*, 28(3):438–451, 2010. 26, 98

[33] C. Gourieroux and R. Sufana. Discrete time Wishart term structure models. *J. Econom. Dynam. Control*, 35(6):815–824, 2011. 26, 98

[34] J. Guyon. Euler scheme and tempered distributions. *Stochastic Process. Appl.*, 116(6):877–904, 2006. 16, 37, 85

[35] S. Heston. A closed-form solution for pptions with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993. 23, 24, 26, 98

[36] J. Jacod, T. G. Kurtz, S. Méléard, and P. Protter. The approximate Euler method for Lévy driven stochastic differential equations. *Ann. Inst. H. Poincaré Probab. Statist.*, 41(3):523–558, 2005. 16, 37, 84

[37] P. Jeganathan. Some aspects of asymptotic theory with applications to time series models. *Econometric Theory*, 11(5):818–887, 1995. Trending multiple time series (New Haven, CT, 1993). 110, 112

[38] B. Jourdain and A. Kohatsu-Higa. *A Review of Recent Results on Approximation of Solutions of Stochastic Differential Equations*, volume 65 of *Progress in Probability.* Springer Basel, 2011. 16, 37, 84

[39] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Applications of Mathematics (New York).* Springer-Verlag, Berlin, 1992. 14, 16, 37, 84

[40] A. Kohatsu-Higa and P. Tankov. Jump-adapted discretization schemes for Lévy-driven SDEs. *Stochastic Process. Appl.*, 120(11):2258–2285, 2010. 13, 16, 37, 84

[41] V. Konakov and S. Menozzi. Weak error for stable driven stochastic differential equations: expansion of the densities. *J. Theoret. Probab.*, 24(2):454–478, 2011. 16, 37, 85

[42] V. Konakov, S. Menozzi, and S. Molchanov. Explicit parametrix and local limit theorems for some degenerate diffusion processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(4):908—923, 2010. 16, 37, 85

[43] S. Kusuoka. Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus. In *Advances in mathematical economics. Vol. 6*, volume 6 of *Adv. Math. Econ.*, pages 69–83. Springer, Tokyo, 2004. 16, 37, 84

[44] S. Kusuoka. Gaussian K-scheme: justification for KLNV method. In *Advances in mathematical economics. Vol. 17*, volume 17 of *Adv. Math. Econ.*, pages 71–120. Springer, Tokyo, 2013. 13, 15, 16, 20, 37, 40, 74, 75, 85, 89

[45] Y. A. Kutoyants. *Statistical inference for ergodic diffusion processes.* Springer Series in Statistics. Springer-Verlag London, Ltd., London, 2004. 26, 98, 102

[46] L. Le Cam. Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. california Publ. Statist.*, 3:37–98, 1960. 110

[47] L. Le Cam and G. L. Yang. *Asymptotics in statistics*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 2000. Some basic concepts. 110, 112

[48] M. Ledoux, I. Nourdin, and G. Peccati. Stein's method, logarithmic Sobolev and transport inequalities. *ArXiv e-prints*, Mar. 2014. 18, 38

[49] J. J. Levin. On the matrix Riccati equation. *Proc. Amer. Math. Soc.*, 10:519–524, 1959. 114

[50] R. S. Liptser and A. N. Shiryaev. *Statistics of Random Processes. I,II.* Applications of Mathematics (New York). Springer-Verlag, Berlin, expanded edition, 2001. Applications, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability. 26, 98

[51] E. Löcherbach and D. Loukianova. On Nummelin splitting for continuous time Harris recurrent Markov processes and application to kernel estimation for multi-dimensional diffusions. *Stochastic Process. Appl.*, 118(8):1301–1321, 2008. 49

[52] H. Luschgy. Local asymptotic mixed normality for semimartingale experiments. *Probab. Theory Related Fields*, 92(2):151–176, 1992. 111

[53] T. Lyons and N. Victoir. Cubature on Wiener space. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 460(2041):169–198, 2004. Stochastic analysis with applications to mathematical finance. 16, 37, 84

[54] E. Mayerhofer. Wishart processes and Wishart Distributions: An affine processes point of view. *ArXiv e-prints*, Jan. 2012. 27, 99, 104, 106, 107, 127

[55] G. Milstein. Weak approximation of solutions of systems of stochastic differential equations. In *Numerical Integration of Stochastic Differential Equations*, volume 313 of *Mathematics and Its Applications*, pages 101–134. Springer Netherlands, 1995. 14, 16, 37, 84

[56] B. Mond and J. E. Pečarić. On matrix convexity of the Moore-Penrose inverse. *Internat. J. Math. Math. Sci.*, 19(4):707–710, 1996. 100, 101

[57] S. Ninomiya and N. Victoir. Weak approximation of stochastic differential equations and application to derivative pricing. *Appl. Math. Finance*, 15(1-2):107–121, 2008. 14, 16, 37, 84, 91

[58] I. Nourdin, G. Peccati, and Y. Swan. Entropy and the fourth moment phenomenon. *J. Funct. Anal.*, 266(5):3170–3207, 2014. 18, 38

[59] I. Nourdin and G. Poly. An invariance principle under the total variation distance. 15 pages, Oct. 2013. 18, 38, 49

[60] E. Nummelin. A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete*, 43(4):309–318, 1978. 49

[61] G. Pagès. Sur quelques algorithmes récursifs pour les probabilités numériques. *ESAIM Probab. Statist.*, 5:141–170 (electronic), 2001. 101

[62] Y. Prokhorov. A local theorem for densities. *Doklady Akad. Nauk SSSR (N.S) in Russian*, 83:797–800, 1952. 20, 36

[63] P. Protter and D. Talay. The Euler scheme for Lévy driven stochastic differential equations. *Ann. Probab.*, 25(1):393–423, 1997. 16, 37, 84

[64] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes, and Martingales. Vol. 2.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Itô calculus, Reprint of the second (1994) edition. 118, 123

[65] T. H. Rydberg. A note on the existence of unique equivalent martingale measures in a markovian setting. *Finance and Stochastics*, 1(3):251–257, 1997. 115

[66] D. W. Stroock. *Probability Theory, an Analytic View.* Cambridge University Press, Cambridge, 1993. 101

[67] D. Talay. Simulation of stochastic differential systems. In P. Krée and W. Wedig, editors, *Probabilistic Methods in Applied Physics*, volume 451 of *Lecture Notes in Physics*, pages 54–96. Springer Berlin Heidelberg, 1995. 14

[68] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509 (1991), 1990. 14, 16, 37, 84

[69] A. Y. Zaĭtsev. Approximation of convolutions of probability distributions by infinitely divisible laws under weakened moment constraints. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 194(Problemy Teorii Veroyatnost. Raspred. 12):79–90, 177–178, 1992. 49